DOCUMENT RESUME

ED 279 116 EC 191 737

AUTHOR	Stake, Robert E.
TITLE	Issues in Research on Evaluation: Improving the Study of Transition Programs for Adolescents with Handicaps.
INSTITUTION	Illinois Univ., Champaign. Secondary Transition Intervention Effectiveness Inst.
SPONS AGENCY	Office of Special Education and Rehabilitative Services (ED), Washington, DC.
PUB DATE	86
CONTRACT	300-85-0160
NOTE	214p.; For related documents, see EC 191 736-746. Some charts contain small print.
PUB TYPE	Collected Works - General (020) -- Information Analyses (070)
EDRS PRICE	MF01/PC09 Plus Postage.
DESCRIPTORS	*Disabilities; *Education Work Relationship; *Evaluation Methods; Intervention; Program Effectiveness; *Program Evaluation; *Research Design; Research Needs; Secondary Education; Standards; *Transitional Programs; Values

ABSTRACT

The report, which is part of a series, contains seven papers concerned with long range improvement of evaluation of transitional programs serving handicapped youth and young adults. Papers have the following titles, authors, and concerns: "Needed Research on Evaluation: Improving the Study of Transition Programs for Adolescents with Handicaps" by L. De Stefano and R. E. Stake (overview of the Transition Institute's mandate and its projects); "Evaluating Effectiveness: Federal Expectations and Local Capabilities" by L. De Stefano (federal evaluation guidelines and local use of evaluation data for program development); "Transition Issues that Affect Research on Evaluation" by A. Halpern (eligibility, service delivery, service outcome issues); "Evaluation, Societal Values, and Transition Programs" by J. Callan (institutional values and the need for researchers to account for the influence of societal values); "Exploring the Role of Professional Standards in Evaluations of Transition Programs: Area of Needed Research" by J. Nowakowski (issues raised by standards proposed by the Joint Committee on Standards for Educational Evaluation); "Epistemic Elements in Evaluation Research" by D. B. Gowin (conceptual analysis, research needed on evaluators, and epistemology); "Needed Research on Optimizing the Structure of Evaluation Designs: The Issue of Flexibility in Evaluation Work" by N. Smith (flexible design structure and accommodation and factors influencing design flexibility); "Situational Context as Influence on Evaluation Design and Use" by R. Stake (importance of context on evaluator analysis, generalization, and evaluation design). A "Bibliography for Evaluative Research on Transition Projects" is appended and includes some 200 citations. (DB)

7266

ED279116

Issues in Research on Evaluation in Transition

Stake

TRANSITION
INSTITUTE
AT ILLINOIS

EC191737

The following principles guide our research related to the education and employment of youth and adults with specialized education, training, employment, and adjustment needs.

- Individuals have a basic right to be educated and to work in the environment that least restricts their right to learn and interact with other students and persons who are not handicapped.
- Individuals with varied abilities, social backgrounds, aptitudes, and learning styles must have equal access and opportunity to engage in education and work, and life-long learning.
- Educational experiences must be planned, delivered, and evaluated based upon the unique abilities, social backgrounds, and learning styles of the individual.
- Agencies, organizations, and individuals from a broad array of disciplines and professional fields must effectively and systematically coordinate their efforts to meet individual education and employment needs.
- Individuals grow and mature throughout their lives requiring varying levels and types of educational and employment support.
- The capability of an individual to obtain and hold meaningful and productive employment is important to the individual's quality of life.
- Parents, advocates, and friends form a vitally important social network that is an instrumental aspect of education, transition to employment, and continuing employment.

The Secondary Transition Intervention Effectiveness Institute is funded through the Office of Special Education Programs, Office of Special Education and Rehabilitative Services, U.S. Department of Education (contract number 300-85-0160).

Project Officer: Dr. Mel Appell

For more information on the Transition Institute at Illinois, please contact:

Dr. Frank R. Rusch, *Director*
College of Education
University of Illinois
110 Education Building
1310 South Sixth Street
Champaign, Illinois 61820
(217) 333-2325

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

Issues in Research on Evaluation: Improving the Study of Transition Programs for Adolescents with Handicaps

Robert E. Stake

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Frank R.
Rusch

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

EC191737

Table of Contents

Page

Needed Research on Evaluation: Improving the Study of Transition Programs -- Lizanne DeStefano and Robert E. Stake 1

Evaluating Effectiveness: Federal Expectations and Local Capabilities -- Lizanne DeStefano 13

Transition Issues that Affect Research on Evaluation -- Andrew Halpern 30

Evaluation, Societal Values, and Transition Programs -- James Callan 57

Exploring the Role of Professional Standards in Evaluation of Transition Programs -- Jeri Nowakowski 70

Epistemic Elements in Evaluation Research -- D. Bob Gowin 113

Needed Research on Optimizing the Structure of Evaluation Designs: The Issue of Flexibility In Evaluation Work -- Nick L. Smith 135

Situational Context as Influence on Evaluation Design and Use -- Robert E. Stake 160

Bibliography for Evaluative Research on Transition Projects 180

NEEDED RESEARCH ON EVALUATION: IMPROVING THE STUDY OF TRANSITION PROGRAMS FOR ADOLESCENTS WITH HANDICAPS

INTRODUCTION

Because people have always evaluated things we should know a lot about evaluation. Because students, schools and curricula have been examined in many ways, we should know an assortment of ways of putting them to the test. Because the policies supporting the transition of handicapped youth from school to work and on to adult responsibility are matters of public concern and investment, we should base those policies on sound data and validated experience. And, in regard to all three statements, we do. Yet program evaluation is far from perfect. It needs improvement, and research is needed to guide that improvement. This report is a selective guide to organizing research on program evaluation.

Within professional education there is a large and complex specialization of evaluation. Much of the technical work grew out of the development of psychological tests, but each of the professional specialties, e.g., teaching, administering, counseling, nursing; has increased staff talent for recognizing high quality and low quality operation. The volume of federal programs in the last three decades brought into emphasis the formalism of evaluation and reporting. The sophistication grew--yet those intimately acquainted with the support and conduct of evaluation know the utility of evaluation reporting frequently has fallen short of expectation.

Since this is true of evaluation of all social action programs, it might be thought that the search for cause of shortfall should be generic, i.e., without attention to the particular kind of program and setting. But it also might be thought that if any improvement in the evaluation of transition programs is going to occur the search for shortfall should be contextual, attendant to the uniqueness of aid to the handicapped adolescent. Our effort has been to examine both.

Our charge was to draw upon philosophers, research methodologists, evaluation practitioners and special educators: together we would examine the evaluation practices most problematic, the severest contentions among rival methodologists and even advocacies driven by undisciplined intuition or self-serving motivation. Together we would identify issues worthy of study and outline research activities for ourselves and our colleagues in special education and evaluation.

From off-campus we brought in:

Robert Brinkerhoff
Bob Gowin
Andrew Halpern
John Pyecha
Charles Reichardt
Nick Smith
Craig Thornton

to work with members of the Transition Institute staff. At a November meeting issues were discussed and a set of topics selected. Six persons volunteered to write on the topics and were told to be attentive to what the other five were writing. Along the way Jeri Nowakowski joined the writing team. Six months later drafts were reviewed at a gathering that also included William Gephart, Robert Henderson, Richard Jaeger, and Les McLean. Members of the group read and reacted to the drafts. Topics became redefined and titles became those listed here in the table of contents. Even a quick look at the titles reveals both the particularity and generality of concerns.

In these papers we examine particularly the issues that arose when a collection of federally funded projects, namely the Office of Special Education and Rehabilitative Services (OSERS) transition/post-secondary projects, were obligated by federal initiative to improve the amount, quality, and use of evaluation data gathered by the projects. The following is an historical/political account of the inception of the OSERS evaluation initiative. Such matters as the discrepancies between federal and local values and purposes of program evaluation are taken up in the individual papers. The abiding concern of course, is the identification of areas where research is necessary to understand and close gaps between these two evaluation stakeholders.

The Background

On December 2, 1983, the United States Congress revised and extended the Education of the Handicapped Law through a series of amendments known collectively as Public Law 98-199. This legislation had a broad focus, dealing with wide ranging aspects of the education of children with handicaps, from establishing the office of Deputy Assistant Secretary of the Office of Special Education and Rehabilitative Services to authorizing grants to states for the removal of architectural barriers. While its predecessor, P.L. 94-142 was explicit in outlining legislation regarding education of a school aged population, P.L. 98-199 contains lengthy sections on early education for a preschool population and secondary/transitional services for youth with handicaps. The law authorized grants for experimental and innovative programs in both areas.

Sections 625 and 626 of P.L. 98-199 authorized a series of grants funding secondary education, post-secondary education, and vocational services for youth with handicaps. The intended purpose of this funding was to discover ways "to strengthen and coordinate education, training, and related services for handicapped youth to assist in the transition process to post-secondary education, vocational training, competitive employment, continuing education, or adult services (Sec. 626). The types of projects which could be funded under this section included demonstration models, exemplary service delivery models, cooperative models between educational agencies and adult service agencies, and research programs; all with the common goal to help special education personnel and "other appropriate persons, including parents, improve the education and related services for handicapped children and youth" (P.L. 98-199, Sec. 641.a). The spirit of this section was clearly to "cast a broad net" i.e., to fund a diverse assortment of projects who were attempting new things in the areas of secondary/transitional service delivery; and then to see "what works" through systematic program evaluation.

The legislation that funded these secondary/transition intervention projects mandated program evaluation, but in keeping with the spirit of discovery and information gathering, the goals of the evaluation were broad (P.L. 98-199, Sec. 618):

1) to assess progress in the implementation of this Act, the impact, and the effectiveness of State and local efforts to provide free, appropriate, public education to all handicapped children and youth; and

2) to provide Congress with information relevant to policy making and provide Federal, State, and local education agencies with information relevant to program management, administration, and effectiveness with respect to such education.

Evaluation was given considerable emphasis in P.L. 98-199, beginning with a call for the Secretary's review of the evaluation section of each grant application and ending with a report of evaluation findings to Congress. "As appropriate, others involved in, or concerned with the education of handicapped children and youth (Sec 618)" were to be informed as well. Recognizing that some assistance would be required to upgrade the amount, quality, and use of evaluation activities, monies were authorized to provide funded projects with technical assistance in the implementation of evaluation design, analysis, and reporting procedures. Additional contracts were authorized to develop procedures for evaluating vocational training, placement, and transitional services for handicapped youth. The legislation authorized funding also to support a thorough and continuing evaluation of the effectiveness of each program assisted under this Act and for findings to be included in OSERS reports to Congress.

In the 1984 funding year, the legislation in Sections 625 and 626 of 98-199 was translated into practice. In 1985 OSERS began an initiative to find "what works" (program, process, interagency agreement) to facilitate the transition of youth with handicaps from high school to work or to postsecondary education. The plan was ultimately to upgrade the amount and quality of the evaluation data collected in approximately 100 OSERS funded secondary/transition projects. These data, collected at the individual project level but aggregated nationally, would serve a number of OSERS' purposes. The data would give an accurate picture of the state of transition services in the United States today. They would identify areas of success and failure, excess and need. OSERS could have this information readily available to inform policy and funding decisions as they arose in Congress. Information on "successful" transition projects would be disseminated to state and local education agencies, universities, and adult service agencies and replication sites would be started, improving service delivery across the country.

Realizing that individual projects might require assistance to obtain the type and quality of data necessary for this initiative, OSERS set up the 5 year, 3 million dollar Secondary/Transition Intervention Effectiveness Institute at the University of Illinois, Urbana-Champaign. The purposes of the Institute are to 1) provide technical assistance on evaluation methods to OSERS-funded secondary and transition projects; 2) to collect and analyze project data related to project characteristics, processes, and outcomes; 3) to review and synthesize relevant literature on transition and evaluation; 4) to foster communication among project personnel; and 5) to conduct a program of research on evaluation. The Institute is under contract to report its research findings and data summaries to OSERS.

At the same time that the Technical Institute and the OSERS transition evaluation initiative were beginning, President Reagan proposed a million dollar cut that would reduce OSERS 1986 funding of secondary and transitional services and post-secondary programs to 1985 levels. Those levels were to continue in 1987 under the administration's proposed budget. The two major pieces of legislation affecting handicapped people--the Education of the Handicapped Act (EHA) and the Rehabilitation Act (RA)--are both scheduled to be reauthorized this year. If this legislation were not reauthorized or extended, discretionary programs funded under these acts (including the secondary and transition projects) would expire in September, 1986.

Madeline Will, Department of Education assistant secretary and director of OSERS, requested that funding for the discretionary programs be extended for a year under the General Education Provisions Act. The extension would give OSERS and advocacy groups time needed for examination of the effects of the 1983 amendments to P.L. 98-199, especially those dealing with transitional of youth with handicaps from high school to postsecondary environments and the world of work. By 1987, Will said, OSERS would be armed with data from evaluation studies such as those conducted by the Transition Institute and would be able to suggest "specific and valid" legislative changes.

Some members of the "Handicapped" Subcommittee of the U.S. Senate, however, felt that OSERS had had ample time to collect data to verify program success. "These programs are up for reauthorization this year and I will not accept a postponement," said Lowell Weicker, R-Conn, chairman of the Subcommittee. He already had begun reauthorization hearings. "You'd better start getting your facts and figures together," Weicker warned Will, "I prefer to do the job with your input, but if you have no recommendations, too bad." (Education of the Handicapped, 12, 4, 1986).

From the exchange it is apparent that OSERS was looking to evaluation efforts to provide valid data on which to base legislative decision-making, as well as a basis for dimunition of the proposed federal budget cuts. OSERS was willing to devote both time and money to the upgrade and aggregation of evaluation data, but it faced time pressures and a not too sympathetic call of Congress to produce "facts and figures". The situation was complicated by the fact that OSERS did not have direct access to the data; they were dependent on another set of players to obtain it.

The Transition Projects, the other set of players, was a loose collective the OSERS funded secondary/transition projects. It was a heterogenous assembly. In 1985-1986 there were approximately 100 projects funded from 12 to 36 months at an average grant of $100,000. The projects shared a common goal: to develop and/or demonstrate linkages and strategies to smoothe the transition of handicapped students from traditional secondary education programs into post-secondary education or work settings. But they targeted and approached their task differently. In their meta- analysis of the projects, Laird Heal and Allen Phelps of the Transition Institute identified seven types of secondary projects:

--those to facilitate transition from high school to work;

--those to facilitate transition from high school to post-high school training programs;

--those to facilitate transition from high school to college;

--those to provide support services to college students;

--those to facilitate the transition from college to work;

--those to develop cooperative (multi-agency) models for transition; and

--those to replicate model programs or disseminate materials for transitional services.

Within each of these types the projects differed with respect to the number and handicapping conditions of students/clients served. About one third of the projects were funded as university projects. Another third operated out of state and community education or rehabilitation facilities. The remaining third of the projects were distributed among public school districts, community colleges, state departments of education or rehabilitation facilities, trade unions, or private, profit making agencies. Projects operated statewide and in local communities. Some were located in towns of less than 10,000. Others were located in the major metropolitan areas of the country.

Although no project listed program evaluation as a primary objective, all projects included a section describing the evaluation of their project in their grant application. (This section was worth 5-10 points in the grant review process.) Across applications, these sections ranged from 5 sentences to 7 pages in length. The guidelines for completing the grant application recommended the collection of "quantifiable data as a measure of student progress". Most projects complied with this request by proposing an evaluation design that included pre- and post administrations of standardized assessment instruments, comparison of employment statistics before and after participation in a program, or ratings of student/client progress by supervisor, teacher, employer, parent, or the student/client him/herself. A small number of projects proposed cost-benefit analyses of their programs. A similar number were interested in looking at the process data concerning the fidelity or degree of implementation of a specific model of transition services.

Given the nature and extent of the proposed evaluation activities, the projects did not allow themselves many resources to do the job. With average annual budgets of $50,000, projects allocated less than 5% ($2,500) for evaluation. Roughly 25% of the projects proposed hiring an outside evaluator (at less than 10% time) to conduct the evaluation. In most projects, evaluation was the responsibility of the project director or his/her designee within the organization. Aside from inclusion in their final report, most projects did not address the issue of dissemination or utilization of evaluation findings in their grant applications.

Most project directors were unaware of the OSERS evaluation initiative until they were summoned to the Transition Institute arranged first annual meeting in Washington, D.C. in November, 1985. Although a 2 1/2 day meeting agenda was devoted to informing the projects of OSERS intent and the roles and services of the Institute, many project directors left that meeting without a clear idea of what implication the new initiative had for their project or what their relationship to the Institute would be. Six months later, this uncertainty remained.

The Need for Research on Evaluation

Through its technical assistance, evaluation, and research program the University of Illinois Transition Institute staff has been addressing this uncertainty. This document is one of a series of reports outlining the nature of the present federally-funded transition effort and, by example and design, illustrating the role of evaluation in gaining understanding of what is at work and what is working.

While designing the Transition Institute, Frank Rusch and Al Phelps saw the need for devoting a small portion of the total efort to long-range improvement of evaluation. Robert Stake was engaged to direct the effort and with James Callan and Mellen Kennedy (and later Annabelle Dryden) assisting, the panel mentioned earlier was put to work. Literature was reviewed and issues studied. The "deliverable" of the first year's effort is this volume.

The topics identified for initial study were the following:

1. Federal vs. local expectations
2. Values
3. Personal and professional standards
4. Flexibility and structure of designs
5. Contextuality of projects
6. Uniqueness of transition

These evolved into the titles listed in the table of contents.

On a number of occasions the panel was reminded that the most frequently cited need was the need for better criterion instruments. Some panelists disagreed, feeling that the shortcomings of transition services were philosophic and organizational more than psychometric. One even said that weaknesses in instrumentation was a needed check to protect the individualization of attention to students and the adaptation capabiity of the individual project. Other panelists recognized that criterion measurement--to indicate both the accomplishment of students and the accomplishment of the staff--was indeed the highest priority for research on evaluation. But they too concluded that differences among projects and among the handicapped required research with a local focus more than national. The consensus opinion was that background research was needed on the topics listed above.

In these papers we have tried to examine the thinking behind the expectations of evaluation. We recognize the considerable change in view among evaluation authorities (See for example Lee Cronbach's "Prudent Aspirations for Social Inquiry") over the last twenty five years, resulting in less confidence in experimental design, less reliance on a comparative point of view, and more attention to political realities. We are aware that we still rely on research methods aimed at theory development when our purposes are recognizing merit, gaining understanding, and improving programs.

We have found it an unusual and exciting opportunity to contemplate a broad program of needed research. But as Nick Smith reminded us, it was perplexing as well. We are not used to it. We are not ready for it. We found it extremely difficult to be explicit about a needed research design or even a well-honed research question. We constantly found our advocacies impeding recognition of our ignorance. We let our thoughts soar, perhaps too often without concern for practicality or other worldly resting place. Here we did not start by examining present practices closely (though in several of the following papers we say we should do so). Had we gone directly to a critical study of our problems, we might have concluded that any research program should grow more directly from confrontation with real and immediate problems, less directly from abstract review, general needs assessment and desire for sweeping change.

It is common for persons not only at different levels within an organization and in different contexts but even for persons side by side to differ in perspect, value, and purpose. The uniqueness of personal point of view remains even in the most straightforward educational program. Local and federal efforts are seen from many perspectives. An evaluation study can be based on several criteria and standards, but it cannot reproduce the complexity of the ways in which the program is beheld, and it seldom has grounds for pre-empting these views. Understanding the complexity of perceiving and valuing is part of the obligation of the evaluator. How can evaluation effectively serve a pluralistic society, to say nothing of a pluralistic profession, pluralistic agencies and pluralistic institutions? Such questions are among those addressed in the papers to follow.

Lizanne DeStefano
Robert E. Stake
August, 1986

EVALUATING EFFECTIVENESS: FEDERAL EXPECTATIONS AND LOCAL CAPABILITIES

Lizanne DeStefano
University of Illinois at Urbana-Champaign

It is commonly accepted that persons at different positions within or surrounding an organization have different values, purposes and levels of understanding regarding an evaluation of that organization. The presence of these differing perspectives raises several questions relating to their impact on evaluation such as: What effects do these differing perspectives have on the nature and quality of evaluation activity and the use of evaluative information? Can an evaluation that is conducted to serve the purposes of persons at one position adequately address the needs of those at another? When there is an extraorganizational catalyst for the evaluation effort such as the funding agency, what constraints are placed on the project's truthfulness? When the funding agency is the federal government, what perceived and actual expectations are held for the type and quality of the evaluation data?

In this paper I examine these and other questions that arise when a particular set of federally funded projects, the Office of Special Education and Rehabilitative Services (OSERS) secondary/transition projects, were obligated by federal initiative to improve the amount, quality, and use of evaluation data gathered by the projects. In the introductory chapter we recounted the historical/political context of the legislation and resultant funding of the secondary/transition projects and the inception of the OSERS evaluation initiative. Building on that foundation, this paper examines the discrepancies that exist between federal and local expectations, values, and purposes of program evaluation and the resulting implications for evaluation practice. Specifically, this paper draws on interviews with OSERS project directors as well as published literature on evaluation to ascertain: a) the nature of federal expectations or needs for evaluation data; and b) the extent of local capabilities to meet these expectations and provide evaluative data. Technological and methodological issues are discussed as well as those dealing with organizational context and structural constraints and requirements. The conclusion of this paper identifies needed research which might lead to the improvement of program evaluation in general and particularly of programs facilitating the transition of youth with handicaps from school to work or postsecondary training.

The Context

The scenario of a federal funding source mandating the conduct and reporting of program evaluation activity has become increasingly more prevalent in education over the last two decades. In recent years, the period of federally commissioned outside evaluations of large scale federally funded programs appears to have given way to the practice of national aggregation of the findings of locally-conducted individual evaluation studies. Since the passage of P.L. 94-142 in the mid-seventies, administrators of special education programs have become accustomed to providing this type of evaluation data to federal and state agencies. These local evaluation studies tend to resemble each other in their design and in the type of data collected. This is due in part to the fact that the language of P.L. 94-142 and the federal and state regulations that have evolved from it are specific with regard to the rights of school age children with handicaps and the responsibility of state and local education agencies to provide a free and appropriate public education to all students. This specificity of the law and its accompanying regulations has resulted in the widespread conduct of program evaluation as a monitoring activity, mandated largely as a measure of individual programs' compliance with federal and state regulations concerning appropriate identification, placement in the least restrictive environment, individualized program planning, due process procedures, and equal access. The types of evaluation data traditionally reported include: headcounts of pupils served (by handicapping condition); documentation of the existence of required services; and statements detailing the procedures used to access them.

Over time this emphasis on compliance or accountability monitoring in special education has come to be misinterpreted as a measurement of program quality and effectiveness. This misinterpretation is so prevalent that from both the local and federal perspectives, a compliant program is usually synonymous with an effective one. In reality, the processes associated with evaluating compliance and effectiveness are quite different. The main purpose of program monitoring is to provide descriptions of the program--clientele served, expenditures of funds, characteristics of program personnel, services rendered, and so on. This type of monitoring is helpful in ascertaining the extent to which an existing social program is meeting legislative requirements and achieving adequate coverage of the population in need. Program monitoring does not tell how effective a program is at achieving its goals nor does it include attempts to systematically assess the direct and indirect effects of the program (Datta & Perloff, 1979).

As was mentioned in the introductory chapter, the language of P.L. 98-199, the law authorizing the funding of the secondary/transition projects, is more experimental and discovery oriented than P.L. 94-142; supporting a diverse array of innovative programs in an extensive search for effective solutions to the problems of transition. The legislation set similar broad goals for the accompanying evaluation activities. In addition to mandating the monitoring of accountability factors such as progress in implementation of the Act, program management, and program administration, the law strongly advocates in favor of evaluation of the impact and effectiveness of local and state delivery efforts with the goal of finding "what works" in the delivery of transition services (Kaufman, 1985).

Federal Expectations

Finding "what works" (program effectiveness) holds somewhat different implications for evaluation design and data gathering than does describing "what's there" (program monitoring). To many, the former implies the use of preset criteria, usually the stated goals of the project set forth in the grant application, to identify and assess the outcomes of a program. Quantitative assessment of the extent to which program objectives are achieved was clearly the evaluative approach favored by OSERS for the first several funding competitions conducted under P.L. 98-199. The following reviewers' criteria for rating the evaluation section of each proposal submitted were excerpted from actual OSERS grant application material from the secondary/transition funding competitions:

> 6. EVALUATION PLAN (0 to 10 points)
>
> Quality of the evaluation plan for the project. Look for information that shows methods of evaluation that are appropriate for the project and, to the extent possible, are objective and produce data that are quantifiable. Also, look for information that shows the quality of the applicant's plan to evaluate

annually: (1) the grantee's progress in achieving the objectives in its approved application, (2) the effectiveness of the project in meeting the purposes of the program, (3) the effect of the project on persons being served by the project...

D. EVALUATION PLAN (Maximum 10 points)

Review the application for information that shows--

(1) The method of evaluation that is appropriate, objective, and produces quantifiable data.

(2) The significant observable, measurable results expected to be achieved during the project period.

EVALUATION PLAN

The evaluation plan should address the following points:

1. Who is to do the evaluation;
2. What data will be collected, how will it (sic) be collected, verified, and stored for ready accessibility;
3. How will data be analyzed and how will the analysis show progress toward and achievement of objectives;
4. State all evaluation criterion (sic) which will be used;
5. To the extent possible, the evaluation methodology is expected to include such elements as client and employee reaction to the program; the extent to which skills have been learned by the clients; and the extent to which job performance has been improved or otherwise changed as a result of the training; or the extent to which clients have been returned to gainful employment or assisted to live independently in the community.

If these excerpts are representative of the federal government's expectations regarding the types of evaluation data that are to be gathered at the local level by the secondary/transition projects, OSERS interest is concentrated upon outcome measures, quantifiable if possible, that indicate the degree to which a program has achieved its stated goals. If the program objectives are centered around process, i.e. implementation, coordination, collaboration, then presumably quantitative process data would be acceptable for federal need, but the more commonly stated goals of service demonstration are associated with outcome variables such as: number of students placed, duration of employment, or student progress in terms of skills gained when the project is involved in direct service to clients, client progress and reaction to the program is also of interest.

This clearer understanding of OSERS expectations for the type of evaluations to be carried out by the secondary/transition projects gives rise to some concerns about their appropriateness and feasibility. As was discussed in the introductory paper, OSERS needs these data for many purposes. The data would give an accurate picture of the state of transition services in the U.S. today. They would identify areas of success and failure, excess and need. Given the proposed budget cuts and the reauthorization of the two major pieces of legislation affecting youth with handicaps--the Education of the Handicapped Act (EHA) and the Rehabilitation Act (RA), OSERS was under extreme pressure to produce data to inform policy and funding decisions as they arose in Congress. Finally, information on successful transition projects would be disseminated to state and local agencies and replication sites would be started, thereby improving service delivery across the country.

Given the experimental nature of the model demonstration projects funded under P.L. 98-199 and the multiplicity of needs for evaluation data that are stated in the introductory chapter, the restriction of evaluation design to an emphasis mainly on outcome variables is surprising. Outcome data may be most appropriate to describe the current state of transition services on a national basis, but they tell little about the characteristics of the projects themselves. When a major goal of the evaluation initiative is to identify and replicate "successful" transition projects, thereby improving service delivery across the country, the absence of interest in implementation and process data is almost inexplicable. The emphasis on the collection of quantitative data is easier to understand in theory though not in practice. If the intent is to aggregate and compare locally produced findings, it is presumably easier to do so with quantitative data than with narrative or qualitative reports. Unfortunately, there are also substantial problems associated with the aggregation of quantitative data. Those who know the pitfalls of trying to compare nonequivalent groups are familiar with the limitations of this plan.

The absence of specific guidelines or monitoring agency to standardize the types of evaluation data to be collected permits creativity in evaluation design on the part of the projects, but exacerbates problems associated with cross-site aggregation of evaluation data. Use of different instruments across projects to measure student and program characteristics make combination or comparison of data problematic. Under-operationalized or differing definitions of handicapping conditions, job success, and skill acquisition make interpretations of aggregated data difficult. If the federal government has the responsibility to evaluate the secondary/transition projects as a whole, then a system for standardizing the collection of certain evaluation data may be desirable to enable the meaningful aggregation of data across sites. This standardized data may suffice as the only requirement to fulfill the federal government's evaluation expectation, but given OSERS desire to acknowledge the individuality of the projects a more desirable alternative may be to request the collection of a small subset of standardized data as an adjunct to the projects' individual evaluation design. To what extent both of these activities could be carried out, given the resources available at the local level and the priorities of the federal government and the projects, is a question that should be researched before such a plan is activated.

There are other concerns with the use of a quantitative approach to measuring change in special populations. The quantitative emphasis reflects a traditional approach to educational evaluation, namely, the use of quantitative student progress information (i.e., achievement test scores) as an indicator of program effectiveness. The narrowness of this approach has been problematic in the evaluation of educational programs for decades (Stake, 1986). When applied to the secondary/transition projects, the utility of this approach is reduced even further by the difficulty of determining levels of "adequate" progress for an exceptional population as well as the lack of reliable and valid instrumentation to measure the client competencies taught by the projects. The long term outcomes of transition such as community employment, increased independence, and greater lifestyle satisfaction are not easily assessed. In some cases the instrumentation is just being developed to quantify some of these outcomes. In other cases the variables seem "unquantifiable." Cost-benefit analysis and other traditional program evaluation techniques require some conceptual rethinking before they can be widely applied to transition. Some work is already being done in that area (Thornton, 1984), much more is needed to make this technology available at the project level.

We could spend many more pages discussing importance and the shortcomings of the outcome oriented, quantitative approach to evaluation favored by OSERS, and we will return to it in the conclusion of this chapter, but for the moment let us turn to the other group of stakeholders in this evaluation initiative, the OSERS funded secondary/transition projects.

Local Perspectives

In the introductory chapter some time was spent describing the 100 OSERS funded secondary/transition projects in the context of the evaluation initiative. In brief, in accordance with the grant application material, all funded projects submitted an evaluation plan prior to funding. These plans varied widely in terms of their degree of detail, the methods proposed, and the amount of resources allocated to the task. In some cases the projects chose to hire an external evaluator, but in the majority of the projects, the task of evaluation fell to the project director and the program staff. Placing the responsibility for program evaluation under the jurisdiction of the project director raises several issues associated with the objectivity, validity, and objectivity of and evaluation planned and conducted in this manner. Of primary concern is the ability of the project director to objectively evaluate his/her own program. The project director believes deeply in the validity and worth of his program. He/she is vitally involved in seeing that the program succeeds. In many cases the project director derives security from his/her identification with the organization and looks to it as a source of personal advancement and recognition (Gurel, 1975). Although the project director may have the most complete understanding of the goals and activities of the project, the extent of his/her personal involvement may limit his/her ability to collect and communicate objective evaluation data. The limitations may be minimized when the evaluation is conducted for internal use in staff development or program improvement; they may be exacerbated in the case of reporting the extent of program success to an external funding agency such as the federal government. Given the prevalence of project directors acting as evaluators of their own projects, the differential outcomes of these scenarios should be examined.

In addition to objectivity, some project directors may not possess the skills and training necessary to conduct a program evaluation. Like the projects they administer, the project directors are a diverse group, representing a variety of disciplines, educational levels, and professional experiences. In a recent article on evaluating employment services, Schalock and Hill (1986) make the observation:

> If program managers were asked whether they were trained to do the tasks they are currently doing the majority would most likely answer, "no." They would add that they were not trained in program monitoring and evaluation, the terms were not in their lexicon 5 years ago, and that program monitoring and evaluation are still not their favorite activity. If one adds the lack of comprehensive data systems and the confusion over operationalizing and quantifying factors that some consider to be "non-quantifiable," it is no wonder why program personnel become frustrated when asked to provide outcome data of their programs' effectiveness, efficiency, and benefit (Boshen, 1984).

Perhaps a key to shortening the distance between federal expectation and local capabilities regarding the evaluation of federally funded program is first to understand the previous program evaluation experience and training of the project personnel as well as what options for consultation and additional training are available to them. This knowledge should be used to inform technical assistance delivery efforts and personnel preparation curriculum development to make project personnel comfortable and capable producers and/or consumers of evaluation findings.

By placing the responsibilities for the design, conduct, and reporting of evaluation activities at the local level, problems of accuracy of reporting, varying levels of evaluation expertise, and lack of systematicity are all but unavoidable. However, local control creates the opportunity for project directors and their staff (with the necessary assistance) to obtain evaluative information that is most meaningful to them; that which can lead directly to program improvement and informed decision making. Unfortunately, this information is usually gained through study of the implementation and processes of a program and the effectiveness of particular program components (Leviton & Hughes, 1981; Smith, in press) and not by assessing overall effectiveness as advocated in the federal guidelines. Given the limited amount of personnel and monetary resources allocated for evaluation, it may be unlikely that a single locally conducted evaluation can meet both local and federal needs to the fullest extent. Under these circumstances, the project director may be put in the position to give priority to either local or federal needs or to superficially address both, leaving one or both users of evaluation information short-changed.

From the point of view of the project directors, the argument in favor of choosing the local perspective is a strong one. Most project directors agree that some degree of program monitoring and evaluation is essential to a program's success. These evaluations are conducted for several reasons, each having a direct impact on the day to day functioning of the project. Most often evaluation was a part of routine accountability reporting to a local overseeing agency. If the project was nearing the end of its OSERS funding, the evaluation was used to justify the continuation of the project by another funding source. In a few projects, process data were collected on program components and used to guide staff development and program improvement activities. Because they are physically close to, and thus familiar with, their local evaluation audiences, the project directors feel that they know what information is of interest to them. The information needs of the federal government are less clear at the project level but seem to imply increased rigor and complexity in data collection and analysis. As a result, several projects who had been conducting local program evaluation internally for the last several years, made the decision to hire external evaluation consultants to handle the OSERS grant evaluation. If such a step replaces internal evaluations, local utility of the evaluation may be sacrificed.

The rewards of effectively responding to local evaluation purposes often have direct and profound impact on the project (and its director). As a direct result of an evaluation, administrators at an overseeing agency may be made aware of the merits of one of its programs, a local funding source may agree to pick up funding at the end of an OSERS grant period, or a politician may lobby for legislation or funding that supports a demonstrably effective program. The rewards of providing evaluation data at the federal level are less clear. Although the federal government's right to these evaluation data was never questioned, the fact that the federal government was not seen as a primary evaluation audience by most of the projects indicates that the accountability link between federal funding and program evaluation is not firmly established in the projects' minds. On the average, the evaluation section of a grant application is worth 5 to 10 points out of 100 in the review process. In most cases, 4 or 5 points were awarded to even the most meager evaluation plans. Evaluation data are to be included in each continuation proposal but continued funding is granted in the absence of these data. Regulations require that the final report issued at the end of program funding include the results of the completed evaluation study, but there is no funding contingent upon the provision of that data and it is often missing.

The value of conducting local evaluations for federal purposes is further threatened by the tendency of projects to overpromise both activities and outcomes when they submit a grant proposal. This inflation at the start of a project immediately invalidates the use of stated project evaluation may not be long enough to fully understand the impact of their intervention.

A Proposed Program of Research

There is a lot of work to be done to narrow the gap between the two consumers of evaluation discussed in this paper, the federal government and local project directors. Needs for different types of evaluative information, unclear rewards for complying with federal evaluation initiatives, questionable feasibility and utility of an emphasis on outcome evaluation, inflated program objectives, and incompatible reporting cycles are all problems which should be investigated in order to fully understand and alleviate the difficulties encountered in federally mandated reporting of local evaluation data.

The proposed program of research should focus around the series of questions presented here:

What data are currently being collected in the projects? How valuable and reliable are they? How costly are they to collect? How are these data being interpreted and how are they used at the local level? How might these data be used at the federal level? How does evaluation impact both positively and negatively on the projects?

What are the expectations of key personnel in OSERS with respect to the type of data to be collected at the project level? What range of data are acceptable to the federal agency? How closely do they match the type and quality of data currently collected for local purposes? How do they see these data being used to affect legislation and decision-making at the federal level? Are there other uses for the data?

How do funding cycles affect the evaluation programs of federally funded projects? Does longitudinal follow up occur after funding ends? What changes in the grant application/monitoring process can ease the funding cycle problem? Are interim results, process data, per cent completion data acceptable?

What are the disincentives for compliance with federal requests for evaluation data? What are some rewards or mechanisms that would increase the willingness of the projects to provide evaluation data? Can project capabilities be increased?

What is the range of evaluation experience and training of the project directors? What are the advantages/disadvantages of having project directors evaluate their own programs? What alternatives are available?

To what extent is overpromising practiced and accepted at local and federal levels? How can innappropriate goals and timelines be amended early in the life of the program to eliminate interference with evaluation? How can overpromising be discouraged in the grant application process? Can evaluation show adequate progress in spite of overpromising?

What types of evaluation questions are of interest to OSERS and the projects? What areas are of common interest? What types of evaluation activities could be used to address these issues? In what areas is evaluation technology insufficient and/or underdeveloped to answer the questions?

What are reasonable allocations of resources to meet the evaluation needs at both local and federal levels? How can projects estimate the cost of their evaluation effort?

The preceding list is quite incomplete. To fully address the multitude of evaluation questions in the area of transition would extend past the life of the Transition Institute. Within the next 5 years however, there is time to prioritize and begin to investigate the most vital of the issues. The projects stand as a willing and rich source of questions and, in many cases, their answers. Evaluation researchers and experts in other fields may be needed to provide answers to some. Together these groups can work to improve evaluation not only for the federally funded transition projects, but for all federally funded programs and the field of evaluation in general.

References

Boschen, K. A. (1984). Issues in evaluating vocational rehabilitation programs. Rehabilitation Psychology, 21, 1, 37-48.

Datta, L. & Perloff, R. (1979). Improving evaluations. Beverly Hills, CA: Sage Publications.

Gurel, L. (1975). The human side of evaluating human service programs: Problems and prospects. In M. Gettentag and E. Struening (Eds.) Handbook of Evaluation Research. Beverly Hills: Sage Publications.

Kaufman, M. (1985). Closing comments. Proceedings of the 1st Annual Project Directors' Meeting. Washington, D.C.

Leviton, L. C. & Hughes, E. F. X. (1981). Research on the utilization of evaluation: A review and synthesis. Evaluation Review, 5, 525-548.

Schalock, R. L. & Hill, M. L. (1986). Evaluating employment services. In Kiernan and Stark (Eds.), Pathways to employment implications for the future, pp. 285-302.

Scriven, M. (1973). Goal-free evaluation. In E. R. House, (Ed.), School evaluation: The politics and process. Berkeley, CA: McCutchan.

Smith, N. (in press). Needed research on optimizing the structure of evaluation designs: The issue of flexibility in evaluation work. In Stake, R. (Ed.).

Staff. (1986, February 19). House seeks extension of EHA, while Senate wants full reauthorization. Education of the Handicapped, p. 3.

Staff. (1986, February 19). Reagan's 1987 budget would set P.L. 94-142 grants at 1985 levels. Education of the Handicapped, p. 1.

Stake, R. (1986). Quieting reform. Champaign, IL: University of Illinois Press.

Thornton, C. (1984). Benefit-cost analysis of social programs. In R. H. Bruininks and C. K. Lakin (Eds.), Living and learning in the least restrictive environment. Baltimore: Paul H. Brookes Publishing Co.

TRANSITION ISSUES THAT AFFECT RESEARCH ON EVALUATION

Andrew S. Halpern
University of Oregon

In many ways, the issues that pertain to research on evaluation are generic, gaining their articulation through explorations in epistemology, philosophy of science, and the technologies of evaluation and research design. Other chapters in this monograph have approached these issues from such generic frames of reference.

In order to do research on evaluation, however, one must also become familiar with the substantive issues that derive from the object of evaluation which, in the present case, is the transition of students with disabilities from school into adult life. The purpose of this paper is to identify and describe these issues, with the goal of providing a contextual framework for the specification of research priorities that other authors have proposed for the evaluation of transition programs. This contextual framework has been operationalized as a set of standards that should be addressed as a precursor of any attempt to do research on evaluation. These standards are derived from an analysis of issues that have been organized under the headings of student or client eligibility, service delivery, and service outcomes. The standards are presented at appropriate points within the text, and also summarized at the end of the chapter in the form of a check list that can be used to help structure the selection of transition programs for any new research on evaluation that is done.

Eligibility Issues

The term "disability" refers to a wide array of conditions that tend to interfere with a person's ability to assume desired social roles. These conditions include learning disabilities, mental retardation, behavior disorders (sometimes described as emotional disturbance), sensory (vision or hearing) impairment, orthopedic impairment, speech impairment, other health impairments, and multiple disabilities. Both type and severity of disability can affect eligibility for services.

Type of disability. Disability labels, such as the ones described above, are legally required in order to justify expenditures for most publicly sponsored programs in special education and adult service agencies. The validity of these labels, however, can often be challenged at two levels: their definitional adequacy and their programmatic relevance.

Within the context of research on evaluation, inadequate definitions of disability are extremely troublesome because they undercut the validity of any findings that may emerge from studies that are done. In essence, when definitions are inadequate, we cannot know how well or consistently they have been operationalized in the selection of samples for evaluation projects.

Two examples may suffice to illustrate this problem. In the area of learning disabilities, both "psychoeducational" and "behavioral" definitions of this condition have been suggested by different professionals in the field. When the same person is evaluated from these differing perspectives, she/he may be identified as learning disabled through one approach but not by the other (Galagan, 1985). In the area of mental retardation, the construct validity of its major defining characteristic, intelligence, has been challenged both legally and conceptually when it was discovered that a disproportionate number of blacks and chicanos have been labeled as mildly mentally retarded and placed in special education classes (Mercer, 1983). As an outcome of litigation, it is now illegal to even use IQ as a diagnostic criterion for special class placement in some states (Galagan, 1985). Both examples raise the spectre of misdiagnosis and misclassification.

Standard No. 1

The types of disabilities represented in evaluation studies of transition programs should be clearly and unambiguously defined.

Predictive relevance is a second issue that has emerged in the literature on diagnostic labelling. At one level, there is no issue. We need labels in order to document expenditures under the current methods of funding, and such labels are typically derived from "structural" assessments of "personal traits". At another level, a substantial body of research literature suggests that there is little predictive validity for many diagnostic labels (Frey, 1984). For example, the evidence is overwhelming that neither IQ nor traditional psychiatric labels (e.g. schizophrenic) are capable of predicting level of community adjustment (Cohen & Anthony, 1984; Halpern, 1984). On the other hand, a "functional" approach to assessment, identifying behavioral skills and deficits, is highly predictive of such outcomes, when appropriate interventions are provided in response to the functional assessment.

Standard No. 2

The identification of people to be included in evaluation studies of transition programs should be based on a functional assessment of disability.

Level of disability. The issue just presented concerning the need to supplant (or at least augment) traditional labels with a functional assessment cuts across both type and level of disability. Another eligibility issue, which relates primarily to severity of disability, is derived from some regulations that govern service delivery. In the vocational rehabilitation program, eligibility depends in part on the counselor's judgment whether or not the potential client's rehabilitation goal is "feasible". This means that the disability is not so severe as to preclude a successful outcome. Severity also enters the picture as a *requirement* for some services, such as income maintenance, medical care, certain residential programs, and on-going case management. The *issue* that emerges with respect to all of these eligibility decisions is the ambiguity that surrounds both the definition and documentation of "severe". Different applications of this criterion will, of course, affect the selection of people to receive services and the outcomes of any evaluative studies that are done.

> *Standard No. 3*
>
> The term "severely disabled" must be defined operationally, whenever this term is used to describe a sample in an evaluation study of transition programs.

Service Delivery Issues

Once eligibility decisions have been reached, a myriad of service delivery issues emerge as programs are adopted in pursuit of transition goals. These issues have been organized into 5 categories: (1) assessment, (2) curriculum, (3) classroom organization, (4) program coordination, and (5) documentation.

Assessment. There are both conceptual and technical issues in assessment that have a potential impact on the diagnosis of disability and, consequently, on the evaluation of transition programs. The conceptual issues include the contrast between functional and structural assessment, between person and environmental assessment, and between process and product assessment. Methodological issues include format limitations on assessment tools that are imposed by disability, and the level of examinee involvement in the assessment process.

The distinction between structural and functional assessment has already been introduced as a diagnostic issue, where it was argued that functional assessment is more relevant than structural assessment as a mechanism for determining eligibility. A very similar argument can be stated forcefully with respect to service delivery. In essence, functional assessment deals with behaviors that are relatively molecular, such as budgeting or home management skills, whereas structural assessment deals with molar constructs that are inferred from assessment performance, such as intelligence or aptitude. Since most instruction is delivered at the molecular level, functional assessment becomes the desirable approach when instructional decisions are being addressed (Galagan, 1985). Within the context of evaluation of transition programs, one must be aware of those instances (which are all too frequent) when structural assessments are used inappropriately to justify instructional decisions.

The role of environmental assessment is very closely related to the role of functional assessment within the service delivery process, in that both are required in order to facilitate good decision-making. The outcome of a functional assessment is a profile of personal strengths and weaknesses along one or more dimensions of behavior that are relevant to the educational or community adjustment objective that is being addressed. In order to make an appropriate decision about service delivery, however, one must also document the environmental demands and resources that are available to solve a given problem. Once this "environmental assessment" has been completed, an attempt can be made to match needs of the person with environmental opportunities (Cohen & Anthony, 1984; Wallace & Larsen, 1978). Given the importance of environmental assessment within a functional approach to decision-making, it is unfortunate to observe that very little work has been done in the development of procedures and instruments that focus on environmental assessment.

> Standard No. 4
>
> Transition plans for each person being served should contain at least 3 components: a clear statement of transition objectives; a functional assessment of strengths and weaknesses of the person; and an environmental assessment of resources and opportunities that will have an impact on the achievement of transition objectives.

Another conceptual issue in the area of diagnosis emerges from the distinction between a product and a process approach to assessment. In the former approach, which is also the most ubiquitous, assessment efforts are concerned only with the products of past learning, and it is assumed that the examinee's prior opportunities to learn have been adequate. Assessment is geared toward measuring what an examinee can do at the present time without assistance, rather than investigating what might be required in the process of instruction in order to enhance the likelihood of future success. It is possible, however, to focus on the process of instruction as an object rather than an assumption of assessment. When this is done, the assessment question is transformed from "How well does a person perform?" to "What resources are required to attain a given level of performance?" (Halpern, 1984; Close & Halpern, 1984). The traditional product approaches to assessment are not well suited to answer this latter question, which is generally posed within the context of a cost-efficiency model of evaluation. Frequently, this question has also been posed within the framework of a zero-reject model of service delivery, where the onus of the eligibility decision is placed on the capacity of the system to provide rather than the capacity of the client to perform.

Standard No. 5

Client assessment procedures should include the administration of "process" measures, whenever the question of cost-efficiency is being raised within the evaluation of a transition program.

Turning next to technical issues that emerge from the assessment process, there are certain disability-related characteristics that must be considered when selecting an assessment instrument. For example, if a norm-referenced instrument is selected for addressing a particular problem, one must ascertain whether or not people with disabilities similar to the person being evaluated have been included in the standardization sample. Testing format can also present problems, as illustrated by the use of instruments requiring high levels of reading with students who have mental retardation. Unless the skill being evaluated *itself* requires reading, such an instrument will confound reading ability with whatever skill is actually of concern, leading to an interpretation that is spuriously low (Halpern, Lehmann, Irvin, & Heiry, 1982).

Standard No. 6

People with disabilities should be included in the development and standardization procedures of any instruments that are used for client evaluation.

Standard No. 7

The reading requirements of a test's format should not exceed that which is required in order to assess the skill or knowledge that is being evaluated.

A focus on functional assessment leads to another technical issue concerning the extent of examinee involvement in the assessment process. This approach to assessment works best when the examinee is thoroughly involved in all states of the assessment process, including interpretation and ultimate decision-making (Vash, 1984). Although questions have been raised concerning the extent of involvement possible when the person being evaluated has limited abilities and/or certain behavior problems, this issue should not be used as an excuse to exclude such people from active involvement in their own evaluation.

Standard No. 8

The individual being evaluated should be actively involved in the process of both designing the evaluation and interpreting the outcomes of assessment.

Curriculum. Assuming that appropriate content has been identified for the instructional programs being provided to adolescents or young adults with disabilities, certain important issues still arise with respect to the design of instruction. These issues can be described in terms of the contrast between a skills versus an activities approach, an inductive versus a deductive approach, and a scripted versus a flexible approach.

The skills/activities issue in curriculum is somewhat analogous to the structural/functional issue in assessment. The skills approach in curriculum organizes instruction around the attainment of somewhat generic skills, which are presumed then to function as basic tools for subsequent problem-solving by the learner. The potential for learning such skills is also viewed as "developmental", with the acquisition of certain skills serving as a foundation for subsequent acquisition of others. One's ability to solve problems is viewed as a function of the level of skill attainment.

Looking at problem-solving from a functional approach, curriculum developers of this persuasion regard the specification of desired activities in a criterion environment as the starting point of instructional design (Wilcox & Bellamy, 1982). Once these problems or outcomes have been identified, instructional procedures are developed and, if necessary, modified until the activity has been performed at an acceptable level of competence. The onus is placed on the teacher to find "some way" of helping the student achieve the desired outcome, rather than on the student "qualifying" for the type of instruction that is being offered.

The maintenance and generalization of whatever is learned through instruction is another major concern in the area of instructional design for people with disabilities. Especially for those with mental retardation or learning disabilities, the question arises as to whether or not students will be able to apply what they have learned during instruction in other settings and at other times.

Both inductive and deductive instructional methods have been developed for enhancing maintenance and generalization (Stokes & Baer, 19977). The inductive approach presents many examples, from representative environments, of opportunities to perform an instructionally targeted skill or activity. If these examples are well chosen to represent the universe of opportunities, it is assumed that the student will eventually be able to perform the skill or activity in situations that have not previously been experienced. The deductive approach is more cognitive in style, teaching a rule for behavior on the assumption that knowing why is as or more important than knowing how, when the goal of instruction is maintenance and generalization. From the perspective of evaluation, students with severe disabilities seem to respond well only to inductive approaches, whereas students with mild disabilities can respond well to either approach, and there is no clear evidence yet concerning which approach works best (Close, Sowers, Halpern, & Bourbeau).

A third curriculum issue is the contrast between a scripted and a flexible approach to the delivery of instruction. In a scripted approach, the sequence and the content of instruction is very tightly designed, including the very words that are spoken by a teacher as instruction is delivered and student errors are corrected. The assumption in back of this position is that learning occurs best under very tightly controlled conditions of instruction. The more flexible approach, on the other hand, presumes that many learning opportunities cannot be engineered in advance, and that the teacher must be able to recognize such opportunities and respond appropriately. In a sense, the issue becomes one of allocating control over the direction of instruction between the teacher and the student (Close et al., 1985).

Standard No. 9

> The instructional design that structures any transition program being evaluated should be clearly identified with respect to the following dichotomies: skill vs. activity orientation; scripted vs. flexible delivery; and inductive vs. deductive approach toward enhancing maintenance and generalization.

Classroom organization. In addition to the issues of instructional design that have just been presented, there are some related issues concerning classroom organization that have an impact on high school transition programs for students with disabilities. These issues pertain to the location(s) in which instruction is provided.

The classical issue in this area has to do with the contrast between special classes, resource room and mainstream classes which are all viable alternatives for students with disabilities. The choice among these alternatives depends largely on the goals and methods of instruction that have been adopted. As a rule of thumb, special classes should be used only for group instruction in those aspects of the curriculum that are completely unique for students with disabilities, and the other options should be used to provide individualized instruction and to facilitate integration of students with disabilities into the regular high school curriculum. From the perspective of evaluation, one must ascertain whether or not there is a proper match between classroom organization and the goals and methods of instruction.

Standard No. 10

Classroom organization should be carefully matched with the goals and methods of instruction.

A more recent issue with respect to classroom organization is concerned with the delivery of instruction in a school building, as contrasted with some other appropriate location in the community. This issue emerges most clearly when the content of the curriculum is focused on vocational training and/or skills in independent living. In such situations, a strong argument can be mustered for "moving the classroom into the community", even though this presents many logistical and administrative difficulties that must be solved (Falvey, 1985).

Standard No. 11

Community-based instruction should be included in any programs that are designed to achieve non-academic transition objectives.

Collaboration. The successful preparation of high school students with disabilities for transition into the community requires extensive collaboration both within the divisions of a school program and between the school and several adult service agencies. Within the school system, the primary divisions are special education, vocational education, and regular academic education, with additional complexity added through the utilization of support services such as counseling and speech therapy.

The adult service agencies that become involved most frequently are vocational rehabilitation, mental health and mental retardation agencies, vocational training facilities, residential programs, and income maintenance programs.

With all this complexity being encountered, it is not surprising that coordination has emerged as a major problem to be solved, both within school programs and between school programs and adult service agencies. In essence, no-one has typically been assigned responsibility for crossing divisional and/or agency boundaries in order to coordinate services that are or might be provided to a given person (Benz & Halpern, in press). As a consequence of this lack of coordination, any given person may experience both duplication of services as well as lack of a needed service. From the service provider's perspective, lack of coordination tends to produce inefficiency at best and chaos at worst when attempts at collaboration are made.

Certain attitudinal barriers have been a historical deterrent to collaboration. Within school programs, vocational education instructors have frequently been unmotivated to accept students with disabilities into their classes. Such instructors were more concerned with dispelling the reputation of vocational education as a "dumping ground" for students who couldn't succeed in the regular academic curriculum. These instructors also believed that encouraging the participation of students with disabilities would enhance the negative reputation of vocational education. Another example of an attitudinal barrier is the perception of some special education teachers that vocational rehabilitation is basically a "paper shuffling" agency, with no truly useful services to provide. Other examples could be cited, but the underlying theme is the same: a lack of desire to collaborate (Halpern & Benz, in press).

Heavy work loads present another reality that can interfere with the possibilities for collaboration. For example, teachers in the regular vocational and academic programs of high schools are often responsible for between 150 and 200 students in 5 or more daily classes. Preparations, paper grading and teaching time to fulfill these responsibilities is more than a full-time job. How can we ask these people to provide additional time and effort to address the needs of students with disabilities? So long as collaboration requires "adding on" of responsibilities, it will be difficult to entice participation from those whose burdens would grow heavier (Halpern & Benz, in press).

Even when a desire to collaborate is present, there are often some administrative problems and issues that tend to interfere. These include deficiencies in information sharing, ambiguity concerning supervisory roles, and some problems that have come to be known as "similar benefits" regulations.

Lack of information sharing has often been identified as a barrier to effective collaboration. For example, public schools rarely inform adult service agencies of the number of students with disabilities who will be leaving school and may require post-school assistance. Young adults with disabilities may have to apply separately to each adult service agency they need, providing the same application information over and over again. The underlying problem is a lack of effective and appropriate procedures for sharing relevant information (Benz & Halpern, in press).

When collaboration has been negotiated between separate agencies, the question of supervision has often emerged. Who is really responsible for developing and implementing a service plan? If this responsibility is shared between professionals from different agencies, who supervises whom? What happens to the student/client as this responsibility is passed around? Difficulties in answering such questions have led to the demise of some collaborative programs, such as the popular work/study program of the 1960's and 1970's, which involved negotiated agreements between public schools and the vocational rehabilitation agency (Halpern, 1974).

A final potential deterrent to collaboration has emerged from the "similar benefits" requirements of many service programs. Briefly, this requirement states that funds shall not be expended on a given service if that service is available from some other source. The idea is to avoid duplication of services and also preserve agency funds for those mandates that are uniquely its own. Unfortunately, it is sometimes unclear whether one agency has more or less responsibility than another for providing a given service. When this occurs, a client can sometimes get lost between the cracks of agency fiscal responsibility (Benz & Halpern, in press).

Standard No. 12

> Procedures for coordinating school and adult agency services should be clearly identified and formally incorporated into each agency's regulations that govern the provision of transition services.

Documentation. The final service delivery area that has potential impact on evaluation is the documentation that is required as people with disabilities receive services both in school and from adult service agencies. Documentation refers both to planning requirements and to closure practices and requirements as people exit from their programs.

The planning documentation requirements that affect evaluation are derived from the major pieces of Federal legislation which govern most school and adult service programs for people with disabilities. A common feature of this legislation is the requirement for producing "individualized program plans" as a precondition for service delivery. These plans must include several components, including documentation of need for services, long range goals and short term objectives, specification of services to be delivered, and timelines for implementation of the plan. All affected parties must participate in the development of the plan, including service providers, service recipients, and family members, when appropriate. The planning process must also be repeated and updated periodically.

The outcome of this process is the development of a formal document called the individualized educational plan (IEP) in special education and the individualized written rehabilitation program (IWRP) in the vocational rehabilitation agency. These documents serve several purposes, including the structuring of a data base that can be used for evaluation. As transition programs and services become more and more prevalent, it is reasonable to anticipate the need for an "individualized transition plan" to document both the process and the outcomes of transition planning.

Standard No. 13

Transition services and desired outcomes should be formally articulated in individualized transition plans for each person receiving services.

The "closure" documents that accompany the termination of services can also present some problems within the context of evaluation. In education, the issue is whether the person leaving school will receive a diploma or a "certificate of completion". The distinction between these two is not always clear and uniform across school districts, and the problem will surely become worse as the current movement toward "excellence in education" accelerates, and graduation requirements become more and more stringent. From the perspective of evaluating transition programs, the problem becomes one of attaching meaning to the attainment of a diploma as an outcome criterion (Halpern & Benz, in press).

A somewhat similar problem exists within the rehabilitation agency where closure status is simply documented with a code number indicating overall success or failure in attaining the rehabilitation goal(s) expressed in the IWRP. Since these goals can vary greatly in their complexity and qualitative merit, a dichotomous "on/off" scoring system will minimize the interpretative power of evaluation studies that utilize such scores.

Standard No. 14

Meaningful exit documents should be developed and used for certifying the extent to which individual transition objectives are achieved.

Service Outcome Issues

Even though eligibility and service delivery issues are obviously relevant to the design and implementation of studies to evaluate transition program, a clear delineation of service outcomes becomes the ultimate concern for any serious probes into the area of summative evaluation. There are several issues of this sort that are particularly relevant to transition programs for students with disabilities. These issues can be characterized as (1) multi-level definition of outcomes, (2) multi-dimensional definition of outcomes, and (3) client versus societal perspectives.

Multi-level definition of outcomes. Differences in severity of disability are likely to have a major impact on the outcomes of transition programs. For example, students with the most severe levels of mental retardation are likely to pursue transition goals that are substantially different from those goals that are sought by students with a single learning disability, such as dyslexia. From the perspective of an outcome level, the issue hinges on whether or not eventual adult adjustment will require ongoing support from other people.

Two current examples of this issue can be cited, one from the area of employment and the other from the area of independent living. A major transition goal for all students with disabilities is to achieve the maximum level of employment that is possible, given any constraints that are imposed by the consequences of disability. For many people, this will mean regular competitive employment in the community. For others, however, competitive employment may only be achievable if it is accompanied by ongoing support from a person whose job is to troubleshoot problems and/or provide the employee with assistance in avoiding problems (Rusch, 1986). Similar programs exist in the residential area, such as "semi-independent living" programs which provide its clients with the minimum amount of training and supervision that is necessary to maintain their homes in the community (Halpern, Close, & Nelson, 1986).

The existence of different outcome "levels" is clearly appropriate for transition programs, but this also presents a problem within the context of evaluation. When we evaluate the impact of transition programs, we must attempt to neither over nor under estimate the "potential" of a given student when outcome goals are selected. If easy (but inappropriate) goals are selected, some programs may appear to perform more effectively than they really do. If difficult (but appropriate) goals are selected, other programs may appear to perform less effectively than those programs that have set their sights at a lower level. Evaluations must be designed very carefully in order to avoid (or at least account for) this source of potential bias.

Standard No. 15

The difficulty of transition goals for individual clients should be documented, taking into consideration the severity of the client's disability.

Multi-dimensional definition of outcome. Current Federal initiatives in the area of transition have defined employment as *the* desired outcome of transition programs (Will, 1984). Community adjustment, however, is obviously a multi-dimensional outcome, including such areas as residential environment and social networks in addition to employment as areas of concern. In order to gain some clarity in this area, there is a need both to define these dimensions operationally, and to examine whatever relationships may exist between the defined dimensions.

Outcome dimensions are not easy to define, because they represent a wide array of social values. Is it more important to earn a lot of money, or to do work that is interesting? Is it more important to have many friends, or to nurture a few deep relationships? Many models of community adjustment have been suggested, and the adoption of a particular model will have obvious implications for how the components of outcome dimensions are operationalized and measured.

Assuming that a multi-dimensional model of community adjustment is eventually adopted as the appropriate goal of transition programs, a new issue will emerge concerning the degree of independence that exists between different dimensions of the model. If the dimensions are highly related to one another, then it is possible that programs geared to address only one dimension will have "spill over" effects on to the others. On the other hand, if the dimensions are largely independent of one another, than programs will need to be designed specifically to attend to each desired dimension. Otherwise, success in one area, such as employment, may be offset by failure in another, such as the quality of social networks (Halpern, Nave, Close, & Nelson, in press).

Standard No. 16

Multi-dimensional goals of transition, including at least employment, residential environment, and social/interpersonal networks, should be incorporated into the transition plans of each person receiving services.

Client versus societal perspective. Assuming the adoption of a multi-dimensional model of transition outcomes, one further issue must be resolved concerning the nature of these dimensions. Most models of adjustment that have been suggested focus on "external" or societal values for their articulation. This includes such areas as employment, independent living, and the development of social networks. The choice of specific elements within such dimensions, however, can vary greatly from individual to individual, to the point where an "internal" or personal definition might vary substantially from an external or societal definition (Halpern, et al., in press). It would therefore be prudent for any transition model of outcome that is adopted to contain both personal and societal components.

Standard No. 17

Client satisfaction should be included as an outcome criterion in any attempt to evaluate the impact of transition services.

Summary

A variety of problems and issues have been reviewed which affect the provision of transition services to adolescents and young adults with disabilities. Out of this analysis has emerged a proposed set of standards. Some standards refer to the methodology of evaluating transition programs which, if ignored, will result in an erroneous interpretation of evaluation outcomes. Other standards are based upon a current analysis of "best practices" within the field of transition programming. Collectively, this set of standards should be viewed as desirable criteria for structuring the selection of programs to be included in research studies on the evaluation of transition programs. This will maximize the power of the research studies by minimizing the confounding effects of poorly designed programs. A check list is included at the end of this chapter to assist in the process of selecting good programs to participate in any research studies that are developed and documenting the potential impact on research outcomes of any standards that are not met by participating programs.

References

Benz, M., & Halpern, A. (in press). A statewide examination of secondary special education for students with mild disabilities: Implications for transition services. Exceptional Children.

Benz, M., & Halpern, A. (1986). Vocational preparation for high school students with mild disabilities: A statewide study of administrator, teacher, and parent perceptions. Career Development for Exceptional Individuals, 9, 3-15.

Close, D. W., & Halpern, A. S. (1984). Measuring the learning potential of severely mentally retarded adults. In A. S. Halpern & M. J. Fuhrer (Eds.), Functional assessment in rehabilitation (pp. 159-170). Baltimore: Paul H. Brookes Publishing Co.

Close, D. W., Sowers, J., Halpern, A. S., & Bourbeau, P. (1985). Programming for the transition to independent living for mildly retarded persons. In K. C. Lakin & R. H. Bruininks (Eds.), Strategies for achieving community integration of developmentally disabled citizens (pp. 161-176). Baltimore: Paul H. Brookes Publishing Co.

Cohen, B. F., & Anthony, W. A. (1984). Functional assessment in psychiatric rehabilitation. In A. S. Halpern & M. J. Fuhrer (Eds.), Functional assessment in rehabilitation (pp. 79-100). Baltimore: Paul H. Brookes Publishing Co.

Falvey, M. (1985). Community-based curriculum: Instructional strategies for students with severe handicaps. Baltimore: Paul H. Brookes Publishing Co.

Frey, W. D. (1984). Functional assessment in the '80s. In A. S. Halpern & M. J. Fuhrer (Eds.), Functional assessment in rehabilitation (pp. 11-43). Baltimore: Paul H. Brookes Publishing Co.

Galagan, J. (1985). Psychoeducational testing: Turn out the lights, the party's over. Exceptional Children, 52, 288-299.

Halpern, A. (1974). Work-study programs for the mentally retarded: An overview. In P. Browning (Ed.), Mental retardation: Rehabilitation and counseling. Springfield, IL: Charles C. Thomas.

Halpern, A. (1984). Functional assesment and mental retardation. In A. Halpern & M. Fuhrer (Eds.), Functional assessment in rehabilitation (pp. 61-78). Baltimore: Paul H. Brookes Publishing Co.

Halpern, A., & Benz, M. (in press). A statewide examination of secondary special education students with mild disabilities: Implications for high school curriculum. Exceptional Children.

Halpern, A., Close, D., & Nelson, D. (1986). On my own: The impact of semi-independent living programs on adults with mental retardation. Baltimore: Paul H. Brookes Publishing Co.

Halpern, A. S., & Fuhrer, J. J. (1984). Functional assessment in rehabilitation. Baltimore: Paul H. Brookes Publishing Co.

Halpern, A. Lehmann, J., Irvin, L., & Heiry, T. (1982). Contemporary assessment for mentally retarded adolescents and adults. Baltimore: University Park Press.

Halpern, A. Nave, G., Close, D., & Nelson, D. (in press). An empirical analysis of the dimensions of community adjustment for adults with mental retardation in semi-independent living programs. Australian and New Zealand Journal of Developmental Disabilities.

Mercer, J. (1983). Labeling the mentally retarded. Berkeley: University of California Press.

Rusch, F. R. (1986). Competitive employment issues and strategies. Baltimore: Paul H. Brookes Publishing Co.

Stokes, T., & Baer, D. (1977). An implicit technology of generalization. Journal of Applied Behavior Analysis, 10, 349-367.

Vash, C. (1984). Evaluation from the client's point of view. In A. Halpern & M. Fuhrer (Eds.), Functional assessment in rehabilitation. Baltimore: Paul H. Brookes Publishing Co.

Wallace, G., & Larsen, S. (1978). Educational assessment of learning problems: Testing for teaching (Chapter 4). Boston: Allyn and Bacon, Inc.

Wilcox, B., & Bellamy, T. (1982). Design of high school programs for severely handicapped students. Baltimore: Paul H. Brookes Publishing Co.

Will, M. (1984). OSERS programming for the transition of youth with disabilities: Bridges for school to working life. Washington, DC: Office of Special Education and Rehabilitative Services.

PROPOSED STANDARDS FOR STRUCTURING RESEARCH ON THE EVALUATION OF TRANSITION PROGRAMS

For each standard listed below, indicate whether or not the standard has been achieved in the transition program being included in the evaluation study. In the space provided at the end of the checklist, describe the anticipated impact of the unmet standard on the design and/or interpretation of the evaluation study being conducted.

Standard No. 1

____yes ____no The types of disabilities represented in evaluation studies of transition programs are clearly and unambiguously defined.

Standard No. 2

____yes ____no The identification of people to be included in evaluation studies of transition programs are based on a functional assessment of disability.

Standard No. 3

____yes ____no The term "severely disabled" is defined operationally, whenever this term is used to describe a sample in an evaluation study of transition programs.

Standard No. 4

____yes ____no Transition plans for each person being served contain at least 3 components: a clear statement of transition objectives; a functional assessment of strengths and weaknesses of the person and an environmental assessment of resources and opportunities that will have an impact on the achievement of transition objectives.

Standard No. 5

____yes ____no Client assessment procedures include the administration of "process" measures, whenever the question of cost-efficiency is being raised within the evaluation of a transition program.

Standard No. 6

____yes ____no People with disabilities are included in the development and standardization procedures of any instruments that are used for client evaluation.

Standard No. 7

____yes ____no The reading requirements of a test's format do not exceed that which is required in order to assess the skill or knowledge that is being evaluated.

Standard No. 8

____yes ____no The individual being evaluated is actively involved in the process of both designing the evaluation and interpreting the outcomes of assessment.

Standard No. 9

____yes ____no The instructional design that structures any transition program being evaluated is clearly identified with respect to the following dichotomies: skill vs. activity orientation scripted vs. flexible delivery and inductive vs. deductive approach toward enhancing maintenance and generalization.

Standard No. 10

____yes ____no Classroom organization is carefully matched with the goals and methods of instruction.

Standard No. 11

____yes ____no Community-based instruction is included in any programs that are designed to achieve non-academic transition objectives.

Standard No. 12

____yes ____no Procedures for coordinating school and adult agency services are clearly identified and formally incorporated into each agency's regulations that govern the provision of transition services.

Standard No. 13

____yes ____no Transition services and desired outcomes are formally articulated in individualized transition plans for each person receiving services.

Standard No. 14

____yes ____no Meaningful exit documents are developed and used for certifying the extent to which individual transition objectives are achieved.

Standard No. 15

____yes ____no The difficulty of transition goals for individual clients is documented, taking into consideration the severity of the client's disability.

Standard No. 16

____yes ____no Multi-dimensional goals of transition, including at least employment, residential environment, and social/interpersonal networks, are incorporated into the transition plans of each person receiving services.

Standard No. 17

____yes ____no Client satisfaction is included as an outcome criterion in any attempt to evaluate the impact of transition services.

ELABORATION OF STANDARDS NOT PRESENTLY MET

Standard No.	Impact of Problem on Proposed Research
__________	______________________________
	______________________________
__________	______________________________
	______________________________
__________	______________________________
	______________________________
__________	______________________________
	______________________________

EVALUATION, SOCIETAL VALUES, AND TRANSITION PROGRAMS

James Callan
St. Patrick's College, Maynooth, Ireland

A central issue for the evaluation of any program and hence for the evaluation research community generally is what is required of us as researchers in order to understand what people value in their programs. Do we know enough about discovering the vision and the trepidation of participants? Do we know how to recognize the socio-political values within the program and in its contexts? This paper deals with the formal inquiry we might undertake to improve our ways of discovering what is valued by sponsors, clients and other stakeholders.

Directly or indirectly, evaluation is a search for what is valued by people. We may be guided by the classical works, e.g., Plato, Spinoza and Bentham and by the more recent writings of the philosophers of ethics, e.g., Moore (1903), Perry (1926), Kohler (1938) and Dewey (1939). What evaluators need to understand about personal valuing has been developed by Apple (1974), Scriven (1974) and Gephart (1981) and is pursued in methodological fashion by Gowin in this publication.

An institutionalization of value commitment is a part of the understanding deserving immediate attention in the continuing professional education of evaluators. Critical values are manifest as much in the routines of institutions as in the behavior of people (Kallos, 1981; Myrdal, 1944). In education, much of the meaning of success and failure of programs is based on values only implicit in institutional contexts.

The search for what is valued in educational programs then requires comprehensive examination of the social and economic structures that support and sustain the schools. Special education transition programs extend in special and complex ways into the workplaces of the community. Often there is a resistance to the spelling out of community corporate and marketplace values. The evaluator's representations of these values perhaps need more than the usual number of validity checks.

Transition from school to work occurs within the complex interconnections between educational, economic and cultural processes.[1] Each school building has its own connections and the system as a whole is something well beyond the sum of the parts (Geertz, 1983). Though appearing quiet and disenfranchised to many an observer, the schools are the seat of dynamic interaction between contentious values across neighborhoods, ethnic groups and generations. Evaluators to transition programs find real challenge in their efforts to connect the internal functions and effects of services preparatory to employment. But often even more difficult is the task of establishing an evidentiary chain between the educational process and economic, cultural, and political forces in the wider society.

The evidence of these relationships is encountered when pursuing many of the issues most troubling to project directors, issues such as:

> To what extent should handicapped youth be integrated into society? Are there ways (beyond education) in which they should occupy a special and partly isolated place?
>
> Should the least handicapped or the most get the greater effort at job preparation?

---

[1] Felice Battaglia (1968) speaks of the contrasting views of work held by Aristotle, da Vinci, Hegel, Marx, Franklin, Adam Smith, and Heidigger, and a "whole line of speculative thought devoted to the prestige of labor as a human value in civilized society and life itself."

What are the limits, if any, to the advisory roles teachers should play in integrating youth into society?

What is the justification for providing subsidized employment to handicapped workers when non-disabled workers are competing for the same jobs?

Each of these policy issues raises value questions. Each invites an examination of decision-rules. Each invites inquiry into the purposes and principles, rationalities and irrationalities, that guide the practices of decision-makers, project directors, teachers, employers, trade union officials, and parents. In such inquiry, the evaluator would be seeking understanding of the basic continuing motivation, purpose and direction of program action. A part of the search is for the project's "sense of character" and "foundational intention." It is a search for what is valued and why.

Such issues move toward fundamental questions of political morality (Young, 1971). Part of the evaluative enterprise is to identify and define issues in such a way that decision makers and other program stakeholders arrive at more sophisticated, if not less troubling, understandings.

INSTITUTIONAL VALUES

Institutional values originate in personal values, become lodged in individual, then group practice, and finally become part of the routine, the rules and regulations, the rites and traditions, of the institution (Meyer, 1980; Sarason, 1971).

In deepfelt ways what do project personnel value?

What value assumptions undergird project activities?

Who holds the determining values?

What is the history of these rules and those discriminations?

Asking such evaluation questions brings into focus the interplay between personal valuing by individual participants and the valuing embedded in institutional practice (House, 1978). Even if we examine, for example the simple question

What are the students learning?

we note that it cannot be neatly separated in our minds from the question

What is the value of what the students are learning?

Our new knowledge, what is being learned this time, is forever colored with our old knowledge, what it is that youngsters are expected to learn in schools. Learning long division is valued. Learning how to take the good gamble is not.

What is learned or experienced is not simply its content. Learning depends on and subsumes the style, the sanctity, of its presentation. Teaching styles are a function of teacher values. Syllabi are an embodiment of community values. Such values are a representation of the program's social context. In transition programs, teaching styles are influenced by the standards for student entry into skilled and semi-skilled jobs. The "moral goodness" of skill is taught. The existing social structures exercise a control and direction on what teachers see as necessary, important, desirable. In large part, such structures determine the character of teaching.

The teaching-learning encounter is part of the wider social enterprise. Encounter and enterprise engage dialectically. In the most of the societies we know the concept of production is vital intellectually and morally. Production attunes us to the wealth of resources and the need for power to extract those resources. In the context of the school and preparation for work, human and material resources limit and shape the curriculum. As Sarason (1971) notes, the schools themselves are cultures and curriculum change is a species of socio-cultural change. If personal power is a vital metaphor in the culture, the need for personal power is taught to those preparing for work: independence, self-determination, assertiveness. These concepts, sometimes named, sometimes not, appear again and again at both institutional and individual levels.

Such considerations suggest that evaluation inquiry be extended into studies that stress culture and the sociology of institutions, but also that it increase cognizance of context and the uniqueness of the site. As Stake emphasizes in another paper in this publication, an educational program occurs in a particular, never greatly generalizable time and place. Schwab (1978) notes that the program comes "with smells, shadows, seats, and conditions outside its walls which may have much to do with what is achieved inside."

Transition programs--and their evaluations--are embedded in a socio-political context. Their evaluation is incomplete if skill level, self esteem, placement, employer satisfaction or even career choice effectiveness is simply the criterion. According to Inglis (1974)

> "...we are talking about the distribution of advantage and disadvantage in society through the differential provision of opportunities to acquire knowledge or to acquire the status that goes with having been exposed to a certain kind of knowledge. There is no question of simply doing it more or less effectively. Effectiveness in such situations only has a meaning when it relates to some set of recognized values or ideals."

What skill level, self esteem, placement, employer satisfaction and career choice mean in the society and its relevant individual subcultures is part of the evaluation responsibility.

Evaluation of transition programs is not only about schools, training and placement, but about social and political policy as well. We poorly understand our capability for attending to the contexts and the policies.

As we inform and discipline ourselves we should study the results to see if our representations of social process improve and become more useful.

AN ILLUSTRATION FROM SPECIAL EDUCATION TRANSITION PROGRAMS

As to handicapped students, what should be taught and how are not problems that can simply be solved by the development and application of theory. The theories include but inevitably fail to explicate their own socialization. They come from people in social, administrative and political structures, reflecting and reacting to the value commitments of those structures.

One rationale for "mainstreaming" is that handicapped students brought more into common classroom experiences will then share more of the culture of the young and its benefits. Notwithstanding "melting pot" rhetoric, the schools hold great store in differentiating students, first to place them in an optimized learning circumstance and later to advise registrars and employers (Spring, 1976). As is apparent in the work of SviShapiro (1980), there is reason to believe that separation through stratification, mainly tracking, pull-out and ability grouping, ensures differentiation that deeply affects a student's school experience. For the handicapped, differentiation signifies to some degree assignment to programs, classes, and assignments of inferior status.

This of course is an ideological as well as empirical perspective. In practice, handicapped students in the mainstreamed classroom are faced with social and academic difficulties more easily avoided in the special classroom. Mainstreaming increases the physical proximity of the disabled to the nondisabled. Some of the deeper sources of student identity, some reflecting the institutional values of the school, are left untouched.

In large part, classroom values derive from the dominant economic and social values in society. Institutional support for and rejection of mainstreaming are not singularly conscious acts on the part of teachers and administrators. They arise out of the basic value assumptions held in society with respect to the distribution of resources, power and cultural experience. Teachers and students do not of course convene in a historical vacuum. Precedents, tradition, and economic "necessity" condition the forms of differentiation and mainstreaming.

In seeking an explanation of the ways mainstreaming is being realized in schools, it is not sufficient for the evaluator to seek out the intentions and meanings which individuals give their actions. The inquiry may need to be supplemented with examinations of the social structures. In the absence of such an analysis, the evaluator is not able to explain comprehensively why individuals behave in the manner they do. Why do they have that attitude toward getting dirty? Why do they use that particular language to describe ability, readiness, failure, and mastery? The reality of mainstreaming is constructed around not only the consciousness of those who participate in the institution, notably the teachers, but also the objective realities of the social and cultural structures within which they reside.

Similar consideration can be given to the emphasis which transition programs place on competitive employment. Federal directives have defined success of projects with respect to the number of handicapped people placed in competitive jobs. Such "employment" is valued in the same way as "mainstreaming," e.g., bringing the handicapped person into the ordinary ways of the society. As with the case of mainstreaming, experience has shown that special education students are often "helped downward." They are trained for and placed on the lowest rung of the occupational ladder, working as unskilled or surplus labor, working well below the average entry level for other young people, not necessarily better than having no job at all. By being introduced into competitive work, yet having highly limited skill, the handicapped person does not share in the ordinary experience of work.

Neither policy nor practice is adequately evaluated in terms of manpower, placement or retention statistics, nor in terms of the personal experience of ordinary youth, nor in terms of personal experience of the handicapped. Key questions in transition program evaluation include those that illuminate the value of competitive work for the handicapped youth who have and do not have it. Researchers should be challenged by the methodological and professional requirements to see how mainstreaming and competitive employment for the handicapped contribute to their well-being and to the well-being of society as a whole.

NEEDED EVALUATION RESEARCH

The writing here is intended to point to the need for evaluators to study their own capabilities for recognizing and recording the societal values which permeate decisions, perspectives and actions of transition programs. Evaluators might arrange circumstances for more careful observation of transition services--aided and challenged by philosophers, sociologists, political scientists, and others from the disciplines, but also from those who have practical experience working with such special populations in the school-to-work context. The research question shifts of course from what teachers, students and employers value to what do evaluators perceive they value.

One way to proceed would be to take the writings of a contemporary commentator such as John Rawls (1971), Charles Murray (1984) or Robert

Bellah (1986), to extract in some depth a key value-construct of theirs (in these cases perhaps justice, paternalism, and individualization, respectively), then to make a critical analysis of an on-going school-to-work program based on this concept, then submitting the analysis to reviewers. An alternative would take the conceptual organizers from issues discussed by evaluation theorists dealing with the social and ethical dimensions of applied social science, writers such as Carol Weiss, Lee Cronbach and Ernest House.

Another research line might develop as an effort to bring the validation procedures know as "triangulation" more directly into an evaluator's recognition of societal values in institutional affairs. For fieldwork generally the concept is discussed by Westie (1957) and Denzin (1970). Data interpretations are made not just in terms of the objectives of the program or rationale for the evaluation but also in terms of competing grand explanations of social process. The idea is not to test these competing theories but to challenge early meanings placed by the evaluation team on observations and ultimately to fortify the interpretations which survive. Empirical research could be undertaken to see if the introduction of social theory to the triangulation process improves the interpretation of school-to-work activities.

In the eyes of evaluators the stereotypic decision-maker seeks clear, unambiguous, focused, legitimated information--even at the risk of oversimplification. The program manager asks the question

What should we do?

Following Reid (1978) the answer may be to contemplate, to deliberate:

> "...practical reasoning is built up through extending, elaborating, and refining the criterion by which actions are to be justified...and made possible by discussion of examples of practice, insights of individuals."

But do evaluation studies which have probed the socio-political structures of transition programs facilitate deliberation or do they foster dissociation and contentiousness? What modes of presentation enhance a necessary toleration of uncertainty without embarrassing the program? Transition programs are rife with uncertainty. As Gautheir (1963) put it,

> "...the sphere of the practical is necessarily the sphere of the uncertain...a problem about what to do...whose final solution is found only in doing something, in acting."

The evaluation community and the transition networks ultimately may profit from empirical studies of the usefulness of decision-makers of evaluations sophisticated in social analysis yet completely honest about the inevitability of uncertainty.

Exploring interactions of societal values in transition programs raises many epistemological, ethical, political and of course methodological problems. Research is needed to support a more insightful and appreciative view on the part of evaluators of the contexts in which handicapped youngsters move from school to work.

References

Apple, M. (1974). The process of ideology and valuing in educational settings. In M. Apple, M. J. Subkoviak, & H. S. Lufler, Jr. (Eds.), Educational evaluation: Analysis and responsibility (pp. 3-34). Berkeley, CA: McCutchan.

Battaglia, F. (1968). Right and good. In P. P. Weiner (Ed.), Dictionary of the history of ideas, volume IV (p. 534). New York: Charles Scribner & Sons.

Bellah, R. N. et al. (1986). Habits of the heart: Individualism and commitment in American life. New York: Harper and Row.

Denzin, N. K. (1970). Sociological Methods. Chicago: Aldine Pub. Co.

Denzin, N. K. (1970). The research act/ A theoretical introduction to sociological methods. Chicago: Aldine Pub. Co.

Dewey, J. (1972). Theory of valuation. Chicago: The University of Chicago Press.

Gauthier, D. P. (1963). Practical reasoning: The structure and foundation of prudential and moral arguments and their exemplification in discourse. Oxford: University Press.

Geertz, C. (1983). Local knowledge. New York: Basic Books.

Gephart, W. J. (1981). Values, valuing and evaluation.

House, E. R. (1978). Assumptions underlying evaluation models. Educational Researcher, 7(3), 4-12.

Inglis, F. (1985). Ideology and the curriculum: The value assumptions of systems builders. Policy Sciences, 18, 3-14.

Kallos, D. (1981). What is studied? Why? How? In T. S. Popkewitz & B. R. Tabachnick (Eds.), The study of schooling. New York: Praeger.

Kohler, W. (1938). The place of value in a world of facts.

Meyer, J. W. (1980). Levels of the educational system and schooling effects. In C. E. Bidwell & D. M. Windham (Eds.), The analysis of educational productivity. Cambridge, MA: Ballinger Publishing.

Moore, G. E. (1903). Principia ethica. Cambridge: At the University Press.

Murray, C. (1984). Losing ground: American social policy 1950-1980. New York: Basic Books.

Myrdal, G. (1944). An American dilemma: The negro problem and modern democracy. New York: Harper & Brothers.

Perry, R. B. (1926). General theory of value. New York: Longmans, Green & Co.

Rawls, J. (1971). A theory of justice. Cambridge, MA: Harvard University Press.

Reid, W. (1978). Thinking about the curriculum: The nature and treatment of curriculum problems. London: Routledge and Kegan Paul.

Sarason, S. (1971). The culture of the school and the problem of change. Boston: Allyn and Bacon.

Schwab, J. (1978). The practical: A language for curriculum. In I. Westbury & N. Wilkof (Eds.), Science, curriculum and liberal education. Chicago: University of Chicago Press.

Scriven, M. (1974). Standards for the evaluation of educational programs and products. In D. Borich (Ed.), Evaluating educational programs and products. Englewood Cliffs, NJ: Educational Technology Publications.

Spring, J. (1976). The sorting machine: National educational policy since 1945. New York: David McKay.

SviShapiro, H. (1980, Summer). Society, ideology and the reform of special education: A study in the limits of educational change. Educational Theory, 30(3), 211-223.

Weiss, C. H. (Ed.). (1977). Using social research in public policy making. Lexington, MA: Lexington Books.

Westie, F. R. (1957, April). Toward closer relations between theory and research: A procedure and an example. American Sociological Review, 149-154.

Young, M. F. D. (Ed.). (1971). Knowledge and control. New directions for the sociology of education. London: Collier-Macmillan.

EXPLORING THE ROLE OF PROFESSIONAL STANDARDS IN EVALUATIONS OF TRANSITION PROGRAMS: AREA OF NEEDED RESEARCH

Jeri Nowakowski
Leadership and Educational Policy Studies
Northern Illinois Universtiy

This paper explores the use of explicit evaluation standards and their ability to improve evaluation practice in general, evaluations of transition programs in particular. To focus the discussion, a specific set of explicit professional standards is considered--the Joint Committee's Standards for the Evaluations of Educational Programs, Projects, and Materials (1981). These standards have been widely circulated but seldom directly used in organizing evaluation studies. Research is needed to clarify the role between formal and informal standards and to determine how explication of Standards can serve evaluators and clients. Explored here also is the possibility that research on the application of the Joint Committee standards might guide the improvement of evaluations of transition programs and other works in special education.

The discussion that follows is divided into five parts. Part I contains a general description of professional standards and standard setting suggesting why both might be of interest to special education. Part II is an overview of the Joint Committee standard setting effort and the standards for evaluation practice they produced. In Part III what it would mean to employ the Joint Committee's Standards in evaluations of transition programs is explored. Part IV contains some specific research-related questions based upon a cursory application of the thirty Joint

Committee Standards to a specific evaluation design for a transition program. In Part V, general research questions of interest when applying the Joint Committee Standards to evaluations of transition programs are anticipated.

I. Professional Standards and Standard Setting

What's to Gain?

Exactly what is a professional standard? From what authoritative base do professional standards spring? And what makes their use more compelling, if indeed it is, than the use of less formalized and more localized standards for evaluation practice?

Professional standards in almost all fields comprise a mixed bag. They contain standard statements both theory and practice based, general and specific, and descriptive and prescriptive. In the field of accounting, for example, a qualitative standard such as Relevance functions like a postulate. Attestation (corroboration) of adherence to such a standard involves a value call. On the other hand, another accounting standard might describe the method for calculating and presenting the net earnings on a per share basis. Attesting to the adherence of a procedural standard is done with infinitely more precision. The professional evaluation standards described here also are a mixed bag. Some standards cut across all evaluation functions (evaluator independence), others are more narrowly focused (report timeliness).

Some standard setters, such as the Joint Committee described in Part II, generate standards within a conceptual organizer. This makes it possible to explore more fully ideological roots and hierarchical distinctions. But in most fields, including accounting, auditing, law and medicine (the so-called self-regulating professions), standards have been generated to meet immediate quality assurance needs and to accommodate emerging research. In these cases, standards grew in patchwork fashion and tend to be rather uneven in specificity.

Professional standards typically are established by boards of persons expert in professional services for use by those providing professional services. (This definition of professional standards does not include standards generated for, but not by professionals, such as those issued by regulating or even licensing agencies.) Professional standard setting boards can be composed only of practicing professionals, but some also include client groups. Professional standards have been generated within professional organizations for the membership (Evaluation Research Society, the AICPA audit standards, and specialty medical standards), and with independent coalitions for a broader professional constituency (the Joint Committee, and financial accounting standards). Professional standards, as referred to here, are intended to guide professionals in their practice, and to define professional services to consumers.

There is variation in what professional groups mean by a standard. A professional evaluation standard as defined by the Joint Committee on Standards for Educational Evaluation (1981), for example, is "a principle commonly agreed to by people engaged in the professional practice of evaluation for the measurement of the value or the quality of an evaluation." The Evaluation Research Society (ERS) defines their standards as "simple admonitory statements" (Rossi, 1982, p. 11). Accounting standards, on the other hand, invoke application of the Rules of Conduct for members of the American Institute of Certified Public Accountants. Departures require justification. Violating professional standards in many fields can bring about professional censure, though voluntary professional organizations have been notoriously ineffective in this area (Ridings, 1980).

So, the authority of professional standards is affected by ideological roots (does authority derive from conventional or normative practice, theory, natural laws, etc.), organizational home (whose standards are they, whose values do they represent), level of specificity (how precisely, if at all, could one measure adherence), and actual use by professionals and society (for instance use for professional preservice education versus adoption by courts for use in litigation). The authority of standards over time depends upon a track record of usefulness, evidence of improved practice, and systematic updating (Ridings, 1980).

Why use explicit professional standards as opposed to adopting or negotiating a less formal set of standards to judge evaluation practice? Because they are explicit, that is to mean straightforward and documented, professional standards allow for front-end agreements about what will constitute good evaluation. Because they are comprehensive they provide an elaborate set of initial considerations for evaluators and clients to winnow down rank. Because they have been negotiated and reviewed over years by practicing evaluators, they collect many sensible and commonly agreed upon principles. Because they provide guideposts for clients and so help them compare alternative evaluation styles and designs. There probably are other reasons, but let us stop with these and then ask "Are there reasons not to use professional standards?"

There are risks associated with the use of professional standards. For example, professional evaluation standards could lead to unnecessary uniformity, might codify conventional practice prematurely, or could limit the conceptualization and growth of evaluation practice. More to the point, use of professional standards might not lead to better practice, and, on the other hand, could result in more costly evaluation. Without studies of the use of professional standards, we don't know how real these problems are, but they are to be considered.

Given the potential risks and benefits, why exactly explore the role of professional standards for evaluations of transition programs? There seems to me to be several good reasons. First, adoption and use of professional standards, even on an exploratory basis, indicates a willingness to be accountable for professional performance. The Joint Committee and Evaluation Research Society standards, along with GAP Audit standards, were incorporated recently into Standards for Program Inspection, Office of Inspector General, U.S. Department of Health and Human Services (April, 1986). There is evidence that federal agencies welcome codified standards, and apparently are willing to use present evaluation standards for their own inspections. This I suppose is in part a political incentive. Additionally, transition evaluation is yet immature in its development, and on occasion narrow in its focus. Both the Joint Committee and Evaluation Research Society standards would broaden systematically the scope of considerations for those involved in transition evaluation. Either would offer a template by which evaluations could be retroactively analyzed or proactively designed. Finally, special education has been a frontrunner in educational evaluation and research in many respects, it seems both appropriate and feasible for this group to produce the first collective set of studies on the effectiveness of professional evaluation standards for one category of programs.

II. The Joint Committee Evaluation Standards

Which Professional Standards?

Standards for the Evaluations of Educational Programs, Projects, and Materials (hereafter referred to as J.C.Standards), were published in 1981 by a Joint Committee representing twelve major educational, research, and psychological organizations.[1] The J.C.Standards were negotiated over a four year period of time and subsequently adopted by their sponsoring organizations. Funded by NIE, NSF, and the Weyerhaeuser Company Foundation, the Joint Committee Standards Project is a unique collaborative effort between organizations whose members use evaluation--client groups--and organizations whose members perform evaluation--practitioner groups. These standards presently are seen as the best available in education, but they will be changed, refined, and perhaps even replaced. It is believed here that they are representative and that research leading to understanding of their use will advance understanding of the use of standards in general.

Several characteristics make the Joint Committee Standards the logical choice in discussions of transition program evaluations: 1) they have been tailored by educational standard setters for educational programs; 2) they represent both client and practitioner perspectives and are constructed for use by both groups; and 3) their aggressive due process procedure lends credibility to their authority and has led to their review and codification by the American National Standards Institute.

The Joint Committee Standards are organized within four domains--Utility, Feasibility, Propriety, and Accuracy (see Part IV for a summary of the standards). Each domain is operationalized through a set of standard statements. For example, it is said that to achieve Utility, audiences must have been identified, the evaluator must be credible and perceived as such, reports must be timely, and so forth. Together the four domains are represented by thirty standards. The presentation of each standard includes an overview, some guidelines for applying the standard, some warnings about misapplication, and a case example of the standard applied in an educational setting.

In an article released at the time the J.C.Standards were published, a research agenda was proposed to study their impact on the field (Ridings & Stufflebeam, 1981). That research agenda is relevant here for two reasons. First, many of the questions and issues proposed in that agenda have not yet been researched. That has left unanswered questions regarding "best practices" in applying the J.C.Standards in a particular context. Secondly, such research questions are applicable to the study of any standards for the evaluation of transition programs.

In Table 1, Ridings and Stufflebeam (1981) identified six major areas of needed research. Briefly, they include: 1) research to determine the degree to which standards reflect or vary from the present state-of-the-art of evaluation practice for a category of studies; 2) study of the value claims inherent in standards and their influence on choice of evaluation procedures and methodology; 3) research on when and how standards are applied during evaluation and how one attests to their application; 4) descriptive studies of the use of misuse of standards; 5) research on the human and material costs of applying standards; and 6) studies of the trade-offs involved in applying individual standards that led to recommendations for their refinement. For the most part, research relative to the Joint Committee standards continues to be needed in all six areas.

III. Evaluation Standards and Transition Programs

How Would Standards be Used?

What might be the role of the J.C.Standards in evaluations of transition programs? Examples of what it means to "use" the J.C.Standards might be helpful; they can be used at each step in the evaluation process. The J.C.Standards provide a set of explicit yardsticks by which the quality of evaluation can be judged. Just as clients sit down with evaluators to articulate program goals (some for which they may be held accountable), the J.C.Standards can provide clients and evaluators a common set of criteria to use in refining evaluation goal statements for which the evaluator should be accountable. At the outset, then, evaluation practice may be rationalized using the J.C.Standards, criteria for its assessment are identified, and responsibility for attesting to the adherence of criteria may be realized. Subsequently, evaluators may use the J.C.Standards to guide evaluation activities, and when modifications are made, to explain to the client what trade-offs are being made in terms of meeting standards. Finally, the J.C.Standards can be used as a meta-evaluation tool, a means by which interim and post hoc evaluations of evaluation can take place by internal or external "auditors."

The initial assumption is that adherence to each standard will optimize the likelihood of sound evaluation practice. In reality there will be trade-offs. Using the J.C.Standards can be like setting out thirty individual prescriptions on the table. In some fashion, client and evaluator line up the standards, ordering them in importance, grouping them for different aspects of the evaluation. If the client wants to move a card, say Report Dissemination, to the bottom of the spread, the evaluator should consider how that will affect adherence to other standards, and the overall quality of the evaluation. The negotiations will be complex, but the *J.C.Standards* would be an explicit and common set of prescriptions on the table. The intention is to promote players playing the same game with both understanding and professional influence over the rules.

Transition Programs as Objects of Evaluations. Transition of handicapped individuals from preschool to school, and through school into adult life, involves the cooperative efforts of many persons and agencies. Rose and Repp (1986) describe transition services as a bridge spanning at least two decades of the handicapped person's life (see Table 2). The bridge is as dependable as its structural supports--those supports include but are not limited to access to high quality preschool programs, responsibly designed elementary and secondary programs, community-based training including vocational experiences, cooperative interagency support, and so forth. Achieving smooth transitions at each step along the way takes communication, planning, coordinating, trained personnel, caring families, and a supportive community structure. As an object of evaluation, transition programs are particularly challenging. No one service provider alone is accountable for transition. To be useful, evaluation should provide feedback to multiple stakeholders serving a target population unique in individual needs.

As a group, evaluation researchers in special education, those responsible for evaluating transition programs, have experience in designing evaluations of integrated delivery systems. This is due in part to the interagency arrangements within which most special education services routinely are delivered. When categorical funding has been obtained, those in the field have had to become experienced in grant preparation, model program designing, and accountability procedures. For over ten years special educators had access to evaluation training through regional workshops produced by the Evaluation Training Consortium--a DDP funded project initially run by Malcolm Provus of the University of Virginia. This training focused on "systems-grounded discrepancy" evaluation. Individual evaluators in special education generally have brought to their practice psychometric and applied behavioral research training. Together these characteristics distinguish this group of evaluation researchers from other evaluators in education. Adoption or application of the J.C.Standards by this group is likely to involve opportunities and constraints unique to both this group of practitioners, and the complex evaluation object--transition programs--to which they direct their efforts. For example, special educators might consider the scope of concerns included in the standards of benefit when seeking and negotiating comprehensive evaluation designs. On the other hand, the generality of standard statements on quantitative analysis may not provide the direction or depth necessary for a field which traditionally has relied heavily on testing for outcome measures.

Use of the J.C.Standards for evaluation of transition programs would be enhanced by initial adoption of evaluators practicing within special education. Given the J.C.Standards client perspective, full use would require endorsement of the J.C.Standards by clients recognizing them to be statements of sound evaluation practice. To the degree that the funding agency is perceived as the major client, this would imply at least unofficial acceptance of the J.C.Standards by that agency. To the degree that clients are defined as service providers, program directors, or handicapped individuals, it would mean involvement of these stakeholding groups, or their representatives, in negotiation of evaluation procedures.

Full and effective use would entail attestation of the standards; that is, reporting on the effectiveness of the evaluation based in part on the degree to which it was conducted in adherence to negotiated and accepted evaluation standards.

IV. Applying the Standards to a Transition Program Evaluation Design

What Issues Would the Standards Raise?

In order to investigate major and unanticipated issues that might result from use of the J.C.Standards for transition evaluations, I examined one proposed design against the J.C.Standards. The evaluation design reviewed is part of a proposal titled "Community Based Transition Project (UCBT) for Severely Handicapped Youth," (CFDA 84.023D) (Hardman & Rose, 1984). This proposal, one of about 200 submitted for funding, was ranked in the top five and funded at a level of about $285,000 for three years. Co-principal investigators were Michael Hardman, Ernest Rose, Clifford J. Drew, Current Project Director John McDonnell, and Barbara Wilcox. Project Director McDonnell and Cliff Drew were principals in designing the evaluation.

Although individual standards were reviewed in relationship only to the evaluation design (the program presently is being implemented), it became clear that elements from each of the thirty standards had been addressed. The most important overall impression was of an appreciable "fit" between the J.C.Standards and the evaluation philosophy and procedures described in this particular evaluation design. To the degree that this evaluation design reflects others in the field, the J.C.Standards would seem to provide a framework which is compatible with implicit and explicit criteria apparently in place.

Summary tables of the transition evaluation design reviewed are included in Appendix A. The evaluation plan is divided into three parts corresponding with a three-pronged program: research and evaluation for the employment training and transition model, evaluation plan for the overall classroom model implementation, and evaluation for personnel preparation (preservice preparation in UCBT) model development. Eight instruments were identified to collect information (see Appendix A); information on their reliability and validity was reported. Both formative and summative purposes were listed, multiple audiences were identified and included as information sources, and evaluation information was planned to be collected and fed back periodically. Multiple measures and multiple information sources were used to optimize validity and reliability of information on a number of variables. While job placements were an important outcome measure, the evaluation plan also included assessment of personal and social adjustment of program recipients. Uses for all information were identified as well as procedures for its collection and dissemination. The overriding impression of the evaluation plan is one of efficient comprehensiveness. It is a system grounded evaluation design, rationalized and relatively fixed, with strong provisions for ongoing management information and accountability indicators.

Even in this cursory application of the J.C.Standards, interesting and researchable issues emerged. Typical issues illustrate the extension of thinking and responsibility the J.C.Standards can provide. Such issues are listed to the right of each of the Joint Committee standards in the section which follows. The left column lists the four domains and thirty standard statements in the Joint Committee Standards.

JOINT COMMITTEE STANDARDS

TYPICAL ISSUES WHEN THE J.C.STANDARDS ARE APPLIED TO EVALUATIONS OF TRANSITION PROGRAMS

A. UTILITY: The Utility Standards are intended to ensure that an evaluation will serve the practical information needs of given audiences.

1.	Audience Identification Audiences involved in or affected by the evaluation should be identified, so that their needs can be addressed	Evaluators are told in this standard to "reach an agreement with the client concerning relative importance of audiences and the information they need." With which client should transition evaluators reach agreement--funders, funded staff, program recipients? Are these audiences likely to have differing needs? How are evaluation audiences' expectations to be ranked? Will the agency funding the program be disproportionately considered?
2.	Evaluator Credibility The persons conducting the evaluation should be both trustworthy and competent to perform the evaluation, so that their findings achieve maximum credibility and acceptance.	Does the credible evaluator (or evaluation team) need to have both substantive (special ed.) training and evaluation research skills? Can projects find and afford such evaluators, or evaluation teams? Will employers and other diverse audiences perceive the same evaluator as credible? The credible evaluator, as defined by this standard, is not biased toward the values of any one audience. Is this a reasonable expectation for transition evaluators?
3.	Information Scope and Selection Information collected should be of such scope and selected in such ways as to address pertinent questions about the object of the evaluation and be responsive to the needs and interests of specified audiences.	This standard calls for evaluation "that assess all important variables (e.g., effectiveness, harmful side effects, costs, response to societal needs, and feasibility)." Are present transition evaluations as broad in scope; can they afford to be? Do fixed designs within funding proposals allow needed flexibility to adjust scope and add information collection strategies? Which client plays the prominent role in adjusting with the evaluator scope and focus throughout the evaluation? How might scope of evaluation information in transition programs be distorted by emphasis on measureable performance data?

4.	Valuational Interpretation The perspectives, procedures, and rationale used to interpret the findings should be carefully described so that the basis for value judgments are clear.	What are the inherent value conflicts in transition programs which carry over into evaluation? For example, are normative data and traditional gain measures inappropriate and inaccurate for this target group? Can evaluators cope with conflicting reports, analyses, and interpretations from different audiences? Is disagreement about what constitutes meaningful outcome adequately depicted? Can evaluators successfully present and interpret value perspectives of key (and perhaps conflicting) audiences?
5.	Report Clarity The evaluation report should describe the object being evaluated and its context, and the purposes, procedures, and findings of the evaluation, so that the audiences will readily understand what was done, what information was obtained, what conclusions were drawn, and what recommendations were made.	Who should evaluators share their information with? What are the ethical or humane considerations to be considered in reporting findings about transition programs? Does preparation of a final, summative evaluation report inhibit production and dissemination of results to other audiences?
6.	Report Dissemination Evaluation findings should be disseminated to clients and other right-to-know audiences, so that they can assess and use the findings.	Do requirements of funded program evaluation reports inhibit their ultimate form, widespread dissemination, and practical use? Is it realistic, as suggested by this standard, to expect that oral and written reports be provided on a regular basis to multiple audiences? How should interim reports and their subsequent impact on the program be audited? How much dissemination should occur outside local and state boundaries? Are reports of transition programs and their evaluations subject to territoriality and proprietary concerns in a competitive grant setting?
7.	Report Timeliness Release of reports should be timely, so that audiences can best use the reported information.	Do transition program evaluators feed back their data to users as soon as they should? Is the criticality of the timeline determined by program needs or by needs of handicapped clients and their parents? If reporting is driven by project funding deadlines, can other internal audiences' information needs be met on a timely basis?

8. Evaluation Impact Evaluations should be planned and conducted in ways that encourage follow-through by members of the audiences.	If transition evaluators attempt to provide follow-through evaluation information for program impact, and if they play a role in that follow-through, have they jeopardized their independence in perceiving, assessing, or reporting program status? Does the transition evaluator play a role in assessing alternative courses of action after the final report is in? Do summative reports written for funders overlook follow-through activities? Are university-based transition evaluators susceptible to preoccupation with "theoretical value of findings at the expense of their application?"

B. FEASIBILITY: The Feasibility Standards are intended to ensure that an evaluation will be realistic, prudent, diplomatic, and frugal.

1. Practical Procedures The evaluation procedures should be practical, so that disruption is kept to a minimum, and that needed information can be obtained.	Do practicable and cost-effective design criteria, as called for by this standard, preclude the selection of research methodology that calls for evaluator independence and controlled settings?
2. Political Viability The evaluation should be planned and conducted with anticipation of the different positions of various interest groups, so that their cooperation may be obtained, and so that possible attempts by any of these groups to curtail evaluation operations or to bias or misapply the results can be averted or counteracted.	When funding client interests are juxtaposed against others within the program's context, what are the conflicts in design, reporting, and interpretation? If the evaluator can't stay "viable" in both settings, what criteria should guide priority setting? What are the advantages and disadvantages of linking evaluation reporting to application for continuation funding?
3. Cost Effectiveness The evaluation should produce information of sufficient values to justify the resources expended.	Should program design and budget subsume evaluation, or should it be an independent component? Is there danger of an evaluation trying to make its information "worth the cost?" To what degree are evaluation costs actually borne by program lines? Are cheap evaluations of transition programs generally as good as expensive ones?

C. PROPRIETY: The Propriety Standards are intended to ensure that an evaluation will be conducted legally, ethically, and with due regard for the welfare of those involved in the evaluation, as well as those affected by its results.

1. Formal Obligation Obligations of the formal parties to an evaluation (what is to be done, how, by whom, when) should be agreed to in writing, so that those parties are obligated to adhere to all conditions of the agreement or formally to renegotiate.	Is the fixed and comprehensive nature of a funding proposal "agreement" a deterrent to meeting new evaluation obligations which emerge during the life of the study?
2. Conflict of Interest Conflict of interest, frequently unavoidable, should be dealt with openly and honestly, so that it does not compromise the evaluation processes and results.	Can evaluators perceived credible in part due to substantive expertise in the area of transition be considered "unbiased?" How does one distinguish professional commitment from bias? Conflict of interest examples identified in this standard include cases in which the evaluation contract might be jeopardized by the report, the evaluator has a particular scholarly or philosophical point which might be advanced by the report, or the evaluator's agency stands to gain or lose based upon the report. How vulnerable is the internal or external transition evaluator in each of these cases?
3. Full & Frank Disclosure Oral and written evaluation reports should be open, direct, and honest in their disclosure of pertinent findings, including the limitations of the evaluation.	What information which, if reported fully, is detrimental to the "ultimate clients" of transition programs? When formative evaluation information is fed back periodically, how should its impact then be considered in the final reports? What are the disclosure problems when the project director and evaluation reporter are the same person?
4. Public's Right to Know The formal parties to an evaluation should respect and assure the public's right to know, within the limits of other related principles and statutes, such as those dealing with public safety and the right to privacy.	Given low-incidence handicap conditions, how can evaluators protect the privacy of individuals and still meet right-to-know standards? In transition evaluation, are three different ethical risks involved in providing evaluation information to right to know audiences?

5. Rights of Human Subjects Evaluations should be designed and conducted, so that the rights and welfare of the human subjects are respected and protected.	Can anonymity rights be upheld in program evaluation reports of small populations? In the case of handicapped individuals (and in light of P.L. 94-142) should parent rights be viewed differently than in cases of regular education? Should parents be considered "subjects" of service as well?
6. Human Interactions Evaluators should respect human dignity and worth in their interactions with other persons associated with an evaluation.	(This standard contains descriptions of ethical and legal considerations for the protection of subjects. This is an area in which special education evaluators have more experience than their regular ed. counterparts. The points raised in this standard appear not to add issues for transition evaluation.)
7. Balanced Reporting The evaluation should be complete and fair in its presentation of strengths and weaknesses of the object under investigation, so that strengths can be built upon and problem areas addressed.	Does emphasis on observables and context aid "balanced" reporting from evaluations embedded within program operations? That is, are evaluations designed to improve and to demonstrate effectiveness also capable of identifying and reporting negative side effects and program weaknesses? Does the formative role of evaluation conflict with this capability? What are the challenges to balanced reporting when professional careers may be built upon the "success" of such projects?
8. Fiscal Responsibility The evaluator's allocation and expenditure of resources should reflect sound accountability procedures and otherwise be prudent and ethically responsible.	Should evaluation expenses be kept independent from program expenses to promote independence? Can a program expect to relinquish funds if new evaluation needs arise? How should evaluation expenditures be audited? How can the right amount to spend on evaluation be determined?

D. ACCURACY: The Accuracy Standards are intended to ensure that an evaluation will reveal and convey technically adequate information about the features of the object being studied that determine its worth or merit.

1. Object Identification The object of the evaluation (program, project, material) should be sufficiently examined, so that the form(s) of the object being considered in the evaluation can be clearly identified.	The interrelated and complex system of components which constitutes a "transition program" are challenging to describe, let alone assess. What can be done to avoid over and under explication of the object to be evaluated?

2. Context Analysis The context in which the program, project, or material exists should be examined in enough detail, so that its likely influences on the object can be identified.	Evaluating contextual events influencing the effectiveness of transition programs is easily a study in and of itself. How does the evaluator balance study of interactions over which service providers have control with study of those events under which they have no control? For example, what is the responsibility of evaluation when a good program design suffers due to shortage of qualified and necessary staff?
3. Described Purposes and Procedures The purposes and procedures of the evaluation should be monitored and described in enough detail, so that they can be identified and assessed.	This standard assumes a unity in purpose and a procedural linearity which may not exist in evaluations of transition programs. Are transition evaluators susceptible to the "pitfall" identified in this standard of assuming that the funding proposal "describes an evaluation's actual purpose and procedures" and further that they "will remain unchanged and be adhered to during the evaluation?" Do the funding report requirements promote discrepancy between purposes and procedures reported and those developing or emerging during the study?
4. Defensible Information Sources The sources of information should be described in enough detail, so that the adequacy of the information can be assessed.	This standard calls for a "variety of information sources," and so may preclude the use of inflexible research design with limited outcome measures. Is such a value structure amenable to to most transition evaluators? Is the resulting design necessarily decision-oriented and formative?
5. Valid Measurement The information-gathering instruments and procedures should be chosen or developed and then implemented in ways that will assure that the interpretation arrived and is valid for the given use.	This standard describes a validation process for all evaluation procedures and instruments. It also recommends the use of multiple measures to assure a full range of variables and to overcome the fallibility of any one single measure. This is a traditional area of concern for special education evaluation research and one in which the evaluation design reviewed was particularly strong.

6. Reliable Information The information-gathering instruments and procedures should be chosen or developed and then implemented in ways that will assure that the information obtained is sufficiently reliable for the intended use.	Reliability issues, another area emphasized in the transition evaluation reviewed, appear typically to be addressed by transition evaluators using multiple assessment instruments and procedures.
7. Systematic Data Control The data collected, processed, and reported in an evaluation should be reviewed and corrected, so that the results of the evaluation will not be flawed.	Do standards promoting practicable and nondisruptive evaluation procedures hamper the collection of quality assurance indicators? Should the project director be in charge of data base? How much process and context information should be scheduled?
8. Analysis of Quantitative Data Quantitative information in an evaluation should be appropriately and systematically analyzed to ensure supportable interpretations.	Which authoritative references and criteria would the transition program evaluator use in adhering to this standard? How would methodological choices be defended?
9. Analysis of Qualitative Data Qualitative information in an evaluation should be appropriately and systematically analyzed to ensure supportable interpretations.	Which authoritative references and criteria would the transition program evaluator use in adhering to this standard? How would methodological choices be defended?
10. Justified Conclusions The conclusions reached in an evaluation should be explicitly justified, so that the audiences can assess them.	Does this standard promote undue adherence to fixed designs and linear evaluation study? It is possible that the formative and client-centered components of an evaluation of a transition program could greatly modify the evaluation questions and even the program design? How are such influences to be interpreted? Does validity increase as specialists in transition are brought in to review interpretations? What project transition conclusions are presently seen as poorly justified?

11. Objective Reporting The evaluation procedures should provide safeguards to protect the evaluation findings and reports against distortion by the personal feelings and biases of any party to the evaluation.	How does the personalization of transition services interfere with objectivity in program evaluation? What are the safeguards for evaluators and clients to guard against bias toward the object? Are they realistic? Is an advocacy study precluded from use by this standard?

V. Building a Research Agenda to Study the Effects of the Joint Committee Standards on Evaluations of Transition Programs

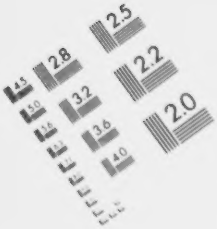
What are the Research Questions?

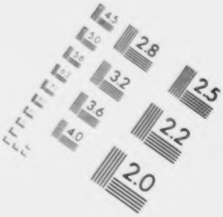
Based upon preceding remarks and assumptions about the Standards, transition programs, evaluators of transition programs, and the initial "fit" between standards and transition evaluations, a more general research agenda is explored below.

Professional Standards vs. Present Practice. To consider how, if at all, use of the Joint Committee Standards might affect evaluations of transition programs, it is important to collect baseline information on the explicit and implicit standards guiding present evaluations of transition programs. This could involve a content analysis of transition evaluation designs to identify and characterize their standards for evaluation performance. We might find, for instance, that transition program evaluations generally adhere to the same technical and ethical principles driving the J.C.Standards. On the other hand, we might find that one domain of the J.C.Standards, such as Utility, is less represented in expectations for evaluation design than another--such as Accuracy. Present standards for good evaluation in this case might translate into evidence of instrument reliability and validity, appropriate sampling, and appropriate analysis procedures. Expectations regarding information scope, valuational interpretation, and audience identification (all standards found in Utility) might be less clear, or even absent.

Research on the explicit criteria presently guiding evaluations of transition would provide three kinds of information. First, use of the J.C.Standards as a template by which transition studies might be

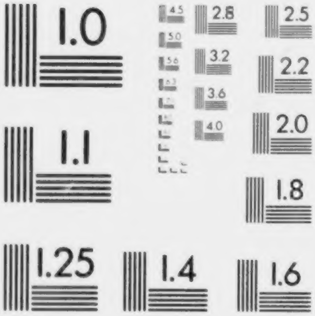
4


2.5
2.8
2.2
3.2
2.0
3.6
4.0
4.5
5.0
5.6
6.3


4.5
5.0
2.8
5.6
3.2
2.5
3.6
2.2
4.0
2.0

PG

100 mm


1.0
4.5
2.8
2.5
5.0
5.6
3.2
2.2
6.3
3.6
2.0
4.0
1.1
1.8
1.25
1.4
1.6

ABCDEFGHIJKLMNOPQRSTUVWXYZ
abcdefghijklmnopqrstuvwxyz1234567890

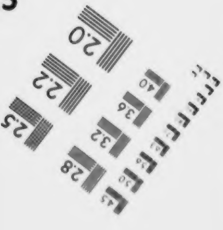
ABCDEFGHIJKLMNOPQRSTUVWXYZ
abcdefghijklmnopqrstuvwxyz
1234567890

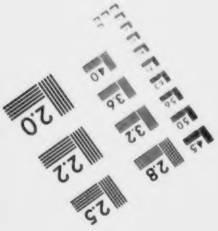
1.0 mm

1.5 mm

2.0 mm

A5


2.0
2.2
4.0
2.5
3.6
3.2
2.8
4.5


4.0
2.0
3.6
3.2
2.2
5.0
2.8
4.5
2.5

described could provide evidence of the relationship between standards for evaluation performance in special education, specifically transition evaluation, and those standards espoused by the Joint Committee. This study would produce as well evidence of the degree to which explicit standards for the hundred or so transition project evaluations vary from or are similar to each other. This analysis would help plot normative or conventional practice, and also would help in identifying exemplary evaluation practice. Secondly, conflicting value claims between the J.C.Standards and present best practices as represented by exemplary transition evaluations could be identified and studied. This would help evaluators anticipate side effects or consequences of applying the J.C.Standards to transition evaluations. Finally, baseline descriptive information on transition evaluations could be compared to other categories of evaluation such as studies of mainstreaming to determine the degree to which "sound evaluation practice" is commonly defined and practiced across special education contexts and evaluation objects.

Because the J.C.Standards present a comprehensive set of concerns, they easily can overburden the small and inexact evaluation design. Ex post facto studies of transition evaluations by researchers using the J.C.Standards must anticipate this problem. The point isn't to determine the percentage of unaddressed standards, but to describe and compare conventional evaluation practice in the field using the standards as a yardstick. The purpose is to identify promising practice and to determine where use of professional standards is likely to improve present substandard practice. Research questions might address whether evaluations which adhered to more of the J.C.Standards, or certain categories of standards, are correlated with high ratings by funders, transition experts in the field, stakeholders, and professional evaluators. And, subsequently, whether successful evaluations as defined by adherence to the J.C.Standards also are associated with successful programs.

Not all standards guiding evaluation practice are explicit, that is documented and direct. Some standards are implicit (suggested, inferred, deducible, understood). Implicit standards for evaluation services are held by clients, audiences, and evaluators. These standards may not be articulated until failure to meet them has occurred. For this reason and others, implicit standards typically are tougher and more slippery customers. What implicit standards are at work in the design, conduct, and evaluation of transition evaluations? One way to investigate is to survey special education evaluators asking them to rank in importance the J.C.Standards. This not-so-clear beginning could be supplemented with case studies of particular evaluators and evaluation studies. Ed Halpern (1983) constructed and applied a protocol based upon principles of auditing which can be used in producing an audit trail of the thinking and valuing taking place at important decision points in a qualitative evaluation. The protocol instructs the "auditor" in how to map the value claims driving evaluator decisions. It would provide interesting evidence regarding the implicit standards guiding the quality of evaluation services in transition evaluations. Moreover, it would allow the field to compare implicit and explicit standards, affording a reality check of sorts.

Client/Evaluator Use of Standards. Because client-evaluator interaction and negotiation play a prominent role in the J.C.Standards, it is important in adhering to the standards that evaluators identify and interact with their client. This is a particular interesting area to research in special education in the case of transition programs. Who are the evaluator's clients in assessing funded transition programs? Is the principal client the federal funding agency? Certainly it is a pivotal audience. Is the client of most concern to the evaluator the directors and administrators of the transition program being evaluated? Indeed, is the evaluator likely to be one of those project staff? Or is the most important client of transition program evaluation the program recipient?

Ernest Rose, Northern Illinois University, has continuing experience with evaluation of transition programs. He believes the majority of evaluators in this area do not consider funders to be their major client. The "ultimate client," he says, is the handicapped individual and the parents of that individual. After them, in importance, fall the service providers. However, he goes on to point out that there is a growing sense of the need for systems analysis and change--that is, evaluation must provide information about and for an interrelated system of people and agencies (conversation with Rose, February 28, 1986). The "ultimate client," from this perspective, is the major character in a cast of hundreds. This makes scope and balance a challenge in transition evaluations. Consider the cross effects of seeking competitive grants and research studies that provide programming for transition subjects. Evaluation reports undermining program effectiveness can undermine program continuation. Which client is of primary concern--the funding feds, the funded staff, or the "ultimate client?" Is it possible fair reporting for one is not fair for the others?

Evaluators and clients need to negotiate which standards will be stressed, and how they will be "measured." Multiple clients complicate this process especially if there are disagreements across clients about what constitutes sound evaluation practice. For example, one client may suggest that utility concerns (such as report dissemination) be compromised for feasibility concerns (such as cost effective and political viability), but another may want the opposite. Case studies of the interaction of multiple clients with the evaluation researcher in transition programs could provide descriptive information about how client needs are actually ranked, and be the basis for guidelines for negotiating evaluation goals among multiple clients.

Interpreting and Guiding Applications of the Standards. Professional standards in any field are based on the assumption that one can distinguish between standard and substandard practice. This assumption is often more easily supported in theory than in practice. When one varies evaluation procedures and methodology, context, scope, and purpose, the variations involved in applying standards make comparative judgments challenging. So, whereas experienced evaluators might readily agree in a crude ranking of three different evaluation studies--they might have more difficulty in agreeing on criteria. It will take multiple observations of the use of standards to guide and evaluate evaluations before the complicated trade- offs and conflicts can be anticipated, and, perhaps resolved. The J.C.Standards may not serve to prescribe sound practice so much as serve as a tool to determine best practice in a specific context.

The fields of accounting and auditing have devised a strategy for guiding and interpreting standards. The strategy takes particular contexts into account and is represented by what is called an "Audit Guide"--a pronouncement demonstrating the application of general standards to a particular setting, e.g., banking. The "Guide" provides specific information and examples for applying generally accepted accounting standards. The "Guide" is a product of a committee composed of those who specialize in accounting within the entity and those who audit accounting. An analogous structure in education might bring together those responsible for designing and delivering transition programs, those responsible for evaluating transition programs, and those responsible for auditing and judging transition programs. Based upon multiple studies of the application of standards to transition evaluations, this group would begin to provide procedural guidelines for how and when the *J.C.Standards* might best be applied for this particular category of study.

Using the *J.C.Standards* to evaluate fifty or more transition program evaluations could help special education evaluators determine how effectively the *J.C.Standards* can serve their needs. Negative side effects such as higher evaluation costs or undo uniformity should be studied. And, conversely, evaluators will want to know if applying the *J.C.Standards* routinely could lead to more effective evaluation procedures and outcomes. A meta- evaluation of transition evaluations also would provide evidence whether other standards, such as the APA Test Standards or the Evaluation Research Society Standards, might better suit their needs.

Methodological Issues in Applying the Standards. From the outset there has been interest in whether application of the *Standards* might influence methodological options for evaluation practice. The *Standards* are by plan methodologically permissive; that is, a decision was made to invite rather than exclude alternative evaluation perspectives, and that decision is reflected in the generality and scope of standard statements. In several research studies, reviewers have examined the congruence of research criteria promoted in the *J.C.Standards* with other sets of explicit standards, for example, the ERS standards and the APA test standards. Most researchers have found substantial overlap between the

ERS and Joint Committee standards (Braskamp & Mayberry, 1982; Cordray, 1982; McKillip & Garberg, 1982; Rossi, 1982; Stockdill, 1984; Stufflebeam, 1982), as well as between the Joint Committee standards and the APA test standards (Bunda, 1982; Impara, 1982; Merwin, 1982; Wardrop, 1982).

Evidence to date suggests that the Joint Committee Standards are compatible with procedures and methodology in other working sets of criteria for inquiry (Wildemuth, 1981). In the case of the APA and Joint Committee standards, the alignment is of consequence to special education. Testing, either for screening, diagnosis, or outcome measures is a critical part of special education measurement strategies and is often embedded in evaluation designs.

Linn (1981) pointed out that the J.C.Standards, by themselves, do not provide the necessary technical detail and support necessary to implement evaluation. According to the Joint Committee, they were not intended to. Limitations in density underline the importance of congruence between the J.C.Standards and other criteria and technical procedures embedded within the evaluation process. Use of the J.C.Standards should not preclude reference to other standards, for example those in testing or quasi-experimental design. There is a precedent for including by reference one set of standards within another as, for example, in the General Accounting Office's reference to the ERS standards within their audit standards. But only as the J.C.Standards are applied can we begin to assess their effect on research methodology.

Also needing study are the methods of using the J.C.Standards. Students in introductory evaluation courses, for example, are inclined to pitch all thirty of the standards at their evaluation design much as if they were two handfuls of beans. Some hit, some don't. But even the more experienced evaluator initially is challenged in using all thirty standards. At what points in an evaluation design standards are applied, how explicitly and systematically, and with what expectations for verification define what it means to "use" the J.C.Standards.

Because most transition programs of interest here are federally funded, stated objectives often must include measureable student performance outcomes. Often the evaluation design is expected to be fixed. Such "front end" planning is commonly prerequisite for funding. Of particular interest, then, is the effect of the J.C.Standards on evaluations designed in conjunction with program designs. For instance, if the J.C.Standards promote more client-evaluator interaction, then their use might not only influence the evaluation design options, but also program design and management. In a decision-oriented and systems-grounded evaluation approach, and such is promoted in the J.C.Standards, formal evaluation is expected to become an integrated, ongoing management tool and less a post hoc, summative assessment. Another area of inquiry, then, might be to study the influence of the J.C.Standards on program design and implementation.

Summary. This paper attempted to review the status of explicit professional standards for evaluation services, and their potential usefulness for evaluations of transition programs in special education. While initial review of the thirty Joint Committee standards indicates that a number of issues would emerge in their use, overall there seems to be interesting possibilities in exploring the use of professional standards for this category of study.

Introduction of explicit standards, particularly the Joint Committee Standards, would provide a number of new and useful research avenues for special education evaluators. Specifically, research using the

J.C.Standards could provide descriptive information on the status of present evaluation practice and the degree to which present standards are compatible with those found in the J.C.Standards. More importantly, adoption of the J.C.Standards to guide evaluation design, evaluation conduct, or meta- evaluation would provide research possibilities in determining effects on the quality and nature of evaluation practice. The research possibilities explored seem promising, the risks appear relatively small, and the payoffs important for the field of special education in general, and for transition programs in particular.

FOOTNOTES

1. Organization members of the Joint Committee in 1980 included: American Association of School Administrators, American Educational Research Association, American Federation of Teachers, American Personnel and Guidance Association, American Psychological Association, Association for Supervision and Curriculum Development, Council for American Private Education, Education Commission of the States, National Association of Elementary School Principals, National Council on Measurement in Education, National Education Association, and National School Boards Association.

2. My special thanks to Ernest Rose, Northern Illinois University for reviewing an initial draft, sharing his work on transition projects and evaluations with me, and improving my understanding of this area. Thanks as well to Alan Repp who also reviewed an initial draft. I am especially grateful for the careful attention and time Bob Stake provided in the development of ideas proposed here. Of course, final responsibility for the contents and limitations of the paper is mine.

References

Braskamp, L. A., & Mayberry, P. W. (1982). A comparison of two sets of standards. Paper presented at the joint annual meeting of the Evaluation Network and Evaluation Research Society, Baltimore, MD.

Bunda, M. (1982). Concerns and techniques in feasibility. Paper presented at the annual meeting of the National Council on Measurement in Education, New York.

Cordray, D. (1982). An assessment of the utility of the ERS standards. In P. H. Rossi (Ed.), Standards for evaluation practice. New directions for program evaluation, No. 15. San Francisco: Jossey-Bass.

Halpern, E. S. (1983, April). Auditing naturalistic inquiries: Some preliminary applications. Paper presented at the annual meeting of the American Educational Research Association, Montreal, Canada.

Hardman, M. L., & Rose, E. (1984). The Utah community-based transition project for severely handicapped youth. Funded by the Office of Special Education and Rehabilitative Services, Department of Education, under CFDA 84.023D.

Impara, J. C. (1982). Measurement and the utility standards. Paper presented at the meeting of the National Council for Measurement in Education, New York.

Joint Committee on Standards for Educational Evaluation (1981). Standards for evaluations of educational programs, projects, and materials. New York: McGraw-Hill.

Linn, M. (1981). Standards for evaluating out-of-school learning. Evaluation News, 2(2), 171-176.

McKillip, J., & Garberg, R. (1982). A further examination of the overlap between ERS and Joint Committee evaluation standards. Unpublished paper. Carbondale: Department of Psychology, Southern Illinois University.

Merwin, J. C. (1982). Measurement and propriety standards. Paper presented at the meeting of the National Council for Measurement in Education, New York.

Ridings, J. M. (1980). Standard setting in accounting and auditing: Considerations for educational evaluation. Unpublished dissertation. Kalamazoo: Western Michigan University.

Ridings-Nowakowski, J. (1982). Standard setting: The crucial issues. A case study of accounting and auditing. Evaluation and Program Planning, 5, 1-9.

Rose, E., & Repp, A. C. (1986). Illinois transition needs assessment survey (ITNAS). Grant application submitted to the Illinois Governor's Council on Developmental Disabilities (CDS-01-86), Springfield.

Rossi, P. H. (Ed). (1982). Standards for evaluation practice. San Francisco: Jossey-Bass.

Stockdill, S. H. (1984, October). The appropriateness of the evaluation standards for business evaluations. Presentation at the Evaluation Network/Evaluation Research Society Joint Meeting, San Francisco.

Stufflebeam, D. L. (1982). An examination of the overlap between ERS and Joint Committee standards. Paper presented at the annual meeting of the Evaluation Network, Baltimore, MD.

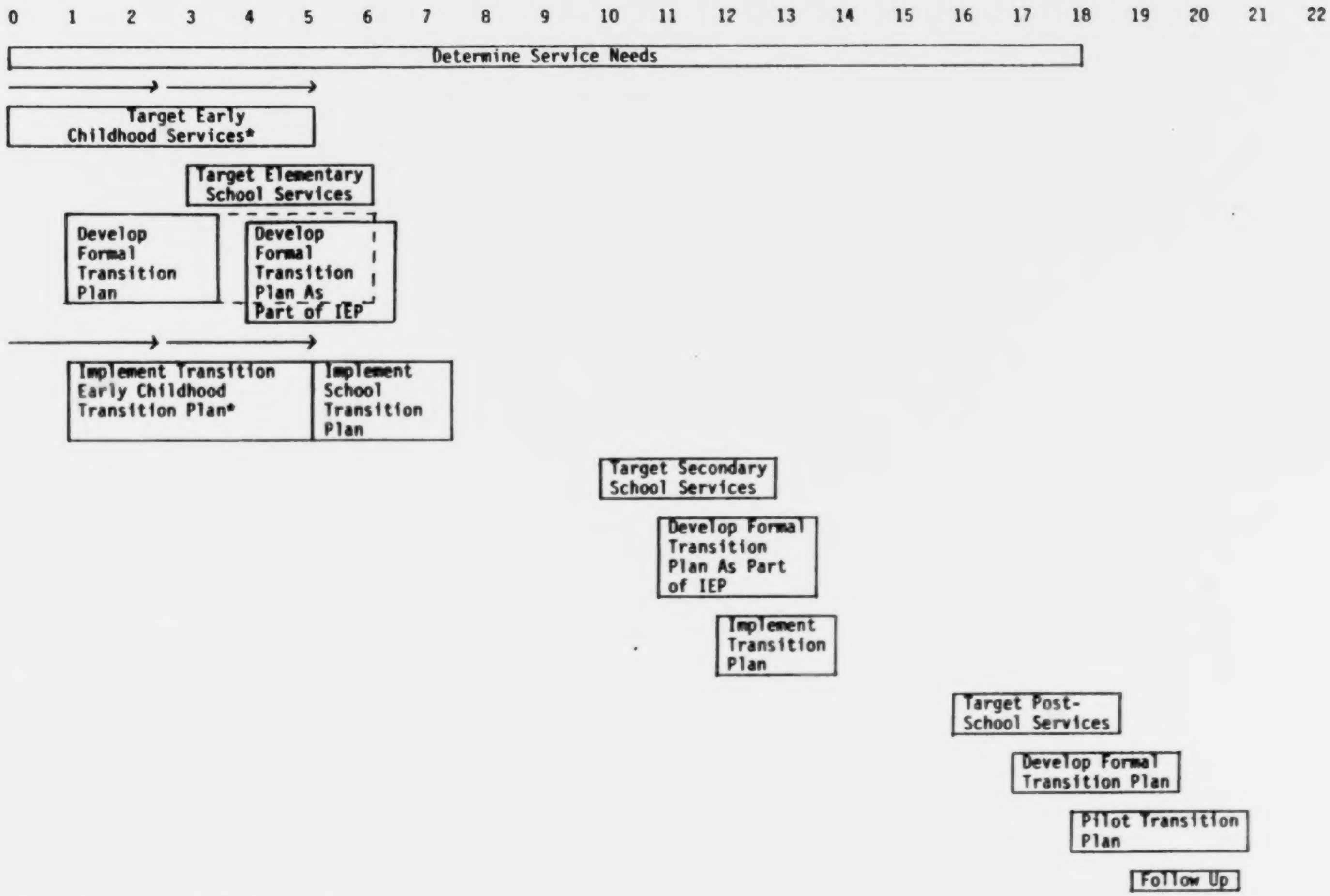
Wardrop, J. C. (1982). Measurement and accuracy standards. Paper presented at the meeting of the National Council for Measurement in Education, New York.

Table 1

Research Agenda to Study the Effects of the Joint Committee Standards

Research Areas	Dependent Variable	Independent Variable	Question of Interest
State of the art	Compliance with *Standards*	Type of evaluation, e.g. formative/summative internal/external large/small Model/design used Stage of evaluation Type of report strategy	What is the present quality of evaluations?
Validity of *Standards*	Applicability of the *Standards*	Program/projects or materials Varying evaluation models Type of evaluation Evaluation outside U.S. Evaluation outside education Adherence to guidelines/ pitfalls/caveats	What are the *Standards*' implicit/explicit claims? Are these claims valid?
Process of applying the *Standards*	Application of the *Standards*	Conflicts in standards Rank ordering of standards Pre-evaluation analysis Evidence of compliance	What are the recurrent problems in applying the *Standards*?
Uses/abuses of the *Standards*	Effectiveness of *Standards*	Evaluation tasks Deciding to evaluate Designing Staffing Managing Analyzing Applying Researching	How are the *Standards* used and with what effect?
Costs/benefits of applying *Standards*	Costs/benefits of *Standards*	Type of evaluation Evaluation context Audience impact Evaluation credibility Number of standards applied	What are the costs and benefits of applying the *Standards*?
Issues within *Standards*	Effectiveness of procedures recommended within *Standards*	Alternative strategies provided within a standard	Is there a "best" strategy which should be indicated in a standard?

Table 2


0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
Determine Service Needs
Target Early Childhood Services*
Target Elementary School Services
Develop Formal Transition Plan
Develop Formal Transition Plan As Part of IEP
Implement Transition Early Childhood Transition Plan*
Implement School Transition Plan
Target Secondary School Services
Develop Formal Transition Plan As Part of IEP
Implement Transition Plan
Target Post-School Services
Develop Formal Transition Plan
Pilot Transition Plan
Follow Up

*Can occur at any point between birth and ages 5 or 6.

Timelines for Transition Planning Activities

Rose, E. (1986). Transition timelines for Oak Park School District 97, Oak Park, IL.

Adapted from McDonnell, J., & Hardman, M. L. (1986). Planning transition of severely handicapped youth from school to adult services: A framework for high school programs. Education and Training of the Mentally Retarded, 20, 275-286.

APPENDIX A

Summary Sheets of UCBT Evaluation Plan

Table 3 -

RESEARCH AND EVALUATION PLAN FOR EMPLOYMENT TRAINING AND TRANSITION MODEL

Issue	Question	Data Source (Instrument)	Measure	Timeline for Collection/Summary	Use of Information
1. Fidelity of Model Implementation	Are all components of the model installed and functioning in the implementation sites. (See Checklist for outcome of various model components.)	Fidelity of Implementation Checklist	-Percent of components meeting standards at the time of program review.	Quarterly	FORMATIVE -Improve implementation support activities -Revise model and Manual -Improve staff training procedures SUMMATIVE -Evaluate applicability of OHS model in classroom settings -Identify program features that appear to affect model utility.
2. Impact of Vocational Model on Classroom Operation.	What types of work training opportunities are available to severely handicapped high school students?	Activities catalog (Work Domain)	-Count of clusters -Specific sites that have/are willing to serve as training sites.		FORMATIVE -Expand catalog entries -Adjust schedule to increase community training time SUMMATIVE -Estimate impact of community based employment training on classroom organization.
	How much time do students spend on work out of the school building?	Classroom Schedule	-Average percent of school week spent in work training.		
3. Impact of Vocational Model on Individual Students	How well do students perform jobs?	Training Packets	Quality and speed prompts	Continuous collection; Summarized Quarterly	FORMATIVE -Adjust training technologies -Adjust the nature of the work training opportunities -Modify transition planning procedures. SUMMATIVE -Evaluate the success of the model on adult life.
	How much out of school work training do students receive in	Individual student resumes	Number of jobs trained Number of job clusters trained Numbers of hours/week		
	What is the employment status of students who graduate from programs using	Follow up questionnaire	Employment status Wages Number of hours/week working.		
4. Social Validity of Model Procedures	How do teachers, administrators, case managers, and employers rate the utility of the model elements?	Model Utility Checklist	-Likert scale satisfaction ratings w/the objectives and the methods of the model.	Quarterly	FORMATIVE -Revision of model elements -Identification of research needs. SUMMATIVE -Evaluate utility of model element.

Table 3
(Continued)

RESEARCH AND EVALUATION PLAN FOR EMPLOYMENT TRAINING AND TRANSITION MODEL

Issue	Question	Data Source (Instrument)	Measure	Timeline for Collection/Summary	Use of Information
5. Social Validity of Model Outcomes	How do teachers, parents, administrators, case managers and employers perceive the model and its results?	model satisfaction scale. Individual student resumes.	-Likert scale satisfaction ratings with model objectives methods, staff and results.	Annually	FORMATIVE -Revise model characteristics -Improve model implementation assistance. SUMMATIVE -Evaluate likelihood of model continuation and adoption in other classrooms.

Table 4

EVALUATION PLAN FOR OVERALL CLASSROOM MODEL IMPLEMENTATION:

Evaluation Issue	Evaluation Question	Data Source (Instrument)	Measure	Timeline for Collection/Summary	Use of Information
1. Fidelity of Model Implementation	Are all components of the OHS model installed and functioning in the implementation sites?	Fidelity of Implementation Checklist	Percent of components functioning at time of observation	Monthly	FORMATIVE -Improve implementation support activities -Revise model and Manual -Improve staff training procedures SUMMATIVE -Evaluate applicability of OHS model in classroom settings -Identify program features that appear to affect model utility
2. Characteristics of Students	What physical, behavioral, and academic disabilities characterize students in the implementation sites?	Student files	Records of significant disabilities and educational events	At beginning of implementation year; at program entry for new student	FORMATIVE -Ensure that students served are severely handicapped SUMMATIVE -Evaluate range of individuals served in classrooms using the model -Estimate restrictions on external validity of model implementation results
3. Student Progress Toward IEP Objectives	Do students in demonstration pilot and implementation classroom achieve objectives specified in IEP's?	Summaries of teacher-collected data	Lists of objectives achieved	Continuous collection; quarterly and annual summary	FORMATIVE -Assist in definition of appropriate objectives -Improve educational interventions SUMMATIVE -Evaluate success of OHS model in achieving IEP objectives for each student
4. Changes in Student Behavior Outside Classroom	Do daily activities of students become more integrated or independent while participating in the OHS classrooms?	Home activity interview	Summary scores for frequency of behaviors indicating community integration and independence	Bi-monthly	FORMATIVE -Assist in selection of objectives that have an impact on students' daily lives -Assist in programing for generalization and maintenance of behavior changes SUMMATIVE -Evaluate success of the model on students' daily lives outside school

Table 4
(Continued)

EVALUATION PLAN FOR OVERALL CLASSROOM MODEL IMPLEMENTATION:

Evaluation Issue	Evaluation Question	Data Source (Instrument)	Measure	Timeline for Collection/Summary	Use of Information
5. Parent Contact and Participation	To what extent to parents of students in the implementation classrooms maintain involvement with the program?	Log of teacher-parent contacts	Frequency of contacts in designated activity categories	Continuous collection with quarterly summary	FORMATIVE -Ensure parent information and involvement in model classroom activities SUMMATIVE -Evaluate success of the model in maintaining parent participation
6. Social Validity of OHS Model Approach and Results	How do parents, administrators, teachers, and other groups affiliated by the model perceive its approach and results?	Model Satisfaction Rating Scale	Likert scale satisfaction ratings with model objectives methods, staff and results.	Annually	FORMATIVE -Revise model characteristics -Improve model implementation assistance SUMMATIVE -Evaluate likelihood of model continuation and adoption in other classrooms
7. Social Validity of Model Elements	How do teachers and other classroom staff rate the utility of model elements?	Model Utility Checklist	Percent of elements rated as useful	Quarterly	FORMATIVE -Revision of model elements -Identification of research needs SUMMATIVE -Evaluate utility of model elements
8. Achievement of Objectives	Are all project objectives met in accordance with projected timelines?	Weekly status review in staff meetings	Status of activities related to each objective	Weekly monitoring; semi-annual summary for SEP and Advisory Committee	FORMATIVE -Improve management activities through reallocation of resources, staff development, etc. SUMMATIVE -Evaluation of project's achievement of projected results
9. Staff Development	Do staff development activities assist in model development, implementation, and evaluation?	Staff development activity record	Number of activities completed by each staff member with a Likert scale rating of relevance to major project activities	Continuous semi-annual summary for SEP and Advisory Committee	FORMATIVE -Assist in planning relevant staff development experiences SUMMATIVE -Evaluate the extent and perceived utility of staff development efforts

Table 4
(Continued)

EVALUATION PLAN FOR OVERALL CLASSROOM MODEL IMPLEMENTATION:

Evaluation Issue	Evaluation Question	Data Source (Instrument)	Measure	Timeline for Collection/Summary	Use of Information
10. Advisory Committee Functioning	Do members of the Advisory Committee participate in scheduled meetings, make recommendations for project improvement, and feel that recommendations have an impact on project operations?	Advisory Committee participant rating	-Attendance -List of recommendations for each project component -Ratings of responsiveness of each component to previous recommendations	Semi-annually, during each meeting of the program review committee	FORMATIVE -Improve quality of project activities by ensuring regular public review -Ensure responsiveness to suggestions of advisory committee SUMMATIVE -Evaluate impact of advisory committee on project activities
11. Extent of Information Dissemination Activities	How broadly disseminated are public information and methodological reports?	Dissemination Log Book	-Number of individuals receiving public and methodological information -Requests for model replication	Continuous collection with semi-annual summary for interim reports	FORMATIVE -Adjust dissemination activities -Estimate required budget for dissemination SUMMATIVE -Evaluate general interest in model activities

Table 5

EVALUATION PLAN FOR PERSONNEL PREPARATION MODEL DEVELOPMENT

Evaluation Issue	Evaluation Question	Data Source	Measure	Data Collection Timeline	Use of Information
1. Incorporation of preservice preparation in UCBT.	Are the elements of UCBT incorporated in the U of U departmental preservice program.	Course syllabus material.	Lists of behavioral objectives for courses on severely handicapped.	Initial incorporation by end of year 2, evaluation of progress on a quarterly basis.	Formative Curriculum decisions and prioritizing of next curriculum development efforts. Summative Year-end assessment of progress in terms of curriculum development changes.
2. Satisfaction/ effectiveness of preservice prep in UCBT.	Are preservice students satisfied with curriculum elements of UCBT.	Course evaluations.	Likert-type assessments completed by students.	Quarterly at the end of each course.	Formative Refinement and revision of course offerings in terms of format and delivery. Summative Year-end assessment of progress in terms of curriculum refinement focusing on format and delivery.
	Are preservice students being effectively trained in the elements of UCBT.	Supervisor/ Employer Evaluations.	Likert-type assessments in questionnaire format completed by supervisors.	Yearly on each graduate.	Formative Refinement and revision of course offerings in terms of content. Summative Year-end assessment of curriculum refinement in terms of content.
	Are preservice students satisfied with the overall prep. received.	Grad. follow-up for preservice students.	Likert-type assessment of satis. with prep. after being in field.	Annual, after student has been on the job for one year.	Curriculum revision decisions based on utility of content and delivery format.
3. Provide inservice training for personnel in the state.	Are the elements of UCBT incorporated in relevant inservice programs.	Course syllabus material.	Lists of behavioral objectives for inservice eval offerings. eation of	Initial incorporation by year 2, progress on a quarterly basis thereafter.	Formative Curriculum decisions and prioritizing of next inservice curriculum development efforts. Summative Year-end assessment of progress in terms of curriculum development/refinement.
4. Satisfaction/ effectiveness of inservice preparation in UCBT.	Are inservice students satisfied with curriculum elements of UCBT.	Course evaluations.	Likert-type assessments completed by students.	Quarterly or at the end of each course.	Formative Refinement and revision of course offerings in terms of format and delivery. Summative Year-end assessment of progress in terms of curriculum refinement focusing on format and delivery.
	Are inservice students being effectively trained in the elements of UCBT.	Supervisor/ employer evaluations.	Likert-type assessments in questionnaire format completed by supervisors.	Yearly on each person trained.	Formative Refinement and revision of course offerings in terms of content. Summative Year-end assessment of curriculum refinement in terms of content.

EPISTEMIC ELEMENTS IN EVALUATION RESEARCH

D. Bob Gowin
Cornell University
(Revised by Bob Stake)

A Transition Project is a political structure with power to intervene in the lives of handicapped persons. Insofar as it is part of Special Education, its power to intervene can be based on principles of educating. The evaluation research required to judge such events is complex. We can follow the advice of A.N. Whitehead to "seek simplicity but preserve complexity." In what follows I make as clear as I can some of the issues I believe need researching. As an educational epistemologist, I am concerned with how human beings construct knowledge about events that educate us -- that help us "get smart."

Events.[1] What events are we concerned with? What is it about a Transition Project that characterizes it as a set of coherent events? (What sorts of events count as "handicapped"?) What is going on? Who knows what is going on? How is knowledge about these events made valid? Can we think of a Transition Project as having a beginning, a middle, and an end? What artifactual regularities, and what natural regularities, are important in evaluation studies in general, and in evaluation studies of "handicapped" in particular? That is, what meaning (significance, value) do these events have?

---

[1] For a different view of events, as seen by a contemporary philosopher (who died recently), see R. M. Martin.

The focus on events is deliberate. As a philosopher, I like to begin with clear cases, then examine some counter-examples, then (inevitably) borderline cases will emerge, and (not so inevitably) imagined cases can serve to test our thinking by going beyond the literal and factual into the imaginary and ideal. So, in short, let's be most careful about our key terms. What, exactly, are we talking about?

Records and Realities.[2] I do not imagine that agreement will be easily expressed. If we already knew what we were talking about, then there would be no need for further research. But one of the sturdy planks of inquiry is the way we make records of events. I give a high valence to records of events -- I call them facts. What are the facts? Who made the facts? (If facts are records of events, then somehow records were made. Often it is a human agency who makes records; sometimes natural events leave reminders that something happened here, e.g., fossils).

Records have great variety: evaluation documents, photographs, audiotapes, videotapes, memos, letters, interviews, budgets-and-bank-balances, calendars, clock-tapes, diaries, authorized signatures, investigative journalism, taped phone calls, wire taps, depositions. (The three rules of lawyers in practice is (a) get in a writing, (b) get it in writing, and (c) get it in writing.) This list of mine is only suggestive, not exhaustive. Needed research here is to open our imaginations about the many, many different ways we use to make records of events. The relation between events and records-of-events constitutes a major grip on making sense of realities. It is our memory. William James concluded that the best way to improve social memory is to improve the way we make records of events. In our time, with such powerful and popular means of silencing, intimate familiarity with a wide variety of records, and finesse with their use, is surely an area of the expansion of research possibilities.

---

[2] See Ulric Neisser, 1982.

Context, Concepts and Meaning.[3] As mystery writers clearly know, the perceived facts to which all agree may turn out to have very different meanings. And as John Dewey was found of saying, there is nothing on the face of a fact that tells you what it means. Evaluators must bring a concept or set of conceptual goggles to bear upon the facts. What concepts do evaluators work with? Concepts, as what we think with, help us to think about facts. Or, put differently, facts are what we think about, not what we think with. Facts, and transformed facts (e.g., data), are largely inert. What supplies us with meaning is the context in which we discuss the relations between events, facts and concepts. All three elements are epistemologically necessary. It is the context that tells us that the term "root" has one meaning in mathematics (square root), another meaning in biology (plant root), and another meaning in social interpretation ("Roots"). As we shift contexts, we shift our intelligence. Put differently, if we change our culture, we change our mind.[4] Needed research here implies culture, context, concepts, meaning - how do these concerns impale on what actions we take to get smart about ourselves and our world? Complexities abound, and too quick precision is too often silencing.

Conceptual Analysis. Philosophers are trained to do concept analysis. I believe that we need philosophic researchers to do a thoughtful analysis of "handicapped." I am moved here to recommend the social philosopher, Michel Foucault (recently killed in a car accident).[5]

---

[3] See Michael Cole, Lois Hood, and Ray McDermott, 1982.

[4] Cf. Michael Coles, 1985.

Foucault's works are not easily summarized, but two major points are of relevance to us. First, he studied the "marginal": prisoners, criminals, sexuality-as-perverse, hospitalized patients, etc. He would have studied the "handicapped", given time. He was concerned about what power in society determines the "deviant," because that also determines the "normal." It is as if he were asking, "What do we do to ourselves when we label someone else deviant?". Is it not the case that all normal people are also handicapped in some ways? Moreover, we can ask a positive question: Do the handicapped have a peculiar gift for recognizing realities the normal overlook? Are they not a treasure, a resource, a valuable aspect to all of human life? The underside of our walkways may have energetic ecologies unrecognized unless studied.

Foucault's second major point: The human sciences, with their increasingly debilitating rigor, are actually misleading us. They mislead us into believing there is a foundational certainty of knowledge. They mislead us into thinking human nature is a natural regularity instead of an artifactual regularity. No underlying set of rails guides human history. Contingency and eventfulness are much more the reality than "scientific determinate causality" suggests. We are more captured by our human sciences than liberated by them. Social power controls knowledge. Methodology and epistemology are secondary considerations in knowledge making.

---

[5] An incisive introduction to Foucault is Mark Philip's piece in the Quentin Skinner edited volume.

We need to continue to investigate this alternative way of thinking about human beings, and the traces and braces of power that control our rationality. At this moment, I feel no better piece of needed research could be undertaken in this context of special education for the handicapped. Who or what controls the meaning that controls research effort?

Research Needed on Evaluators. A line of research I would like to see emerge is the relation between "epistemological commitment" and "evaluation research practice." We have some daunting evidence at Cornell University about biology professors, their teaching assistants, and their expectations about student learning. The evidence so far suggests two puzzling matters: (a) that there exists an unbridgeable, un-negotiable gap between two kinds of epistemological commitments -- i.e., that between those scientists who believe "knowledge is discovered" and those who believe "knowledge is a human construction." This difference is pretty well-known in American society because of the influence of Logical Positivism, Logical Empiricism, Behaviorism, and Scientific Realism (Putnam, Boyd, etc.).

Richard Roty and Richard Bernstein have explicated the pernicious effects of an unforgiving Foundationalism.[6] What we at Cornell have found through interviews with faculty members is that the foundationalists and the constructionists do not communicate with each other. People with different epistemologies have different social concerns. I will discuss foundationalism further in the next section.

---

[6] The two Richards, Rorty, 1980 and Bernstein, 1983, argue against Foundationalism (as well as relativism).

The second puzzling matter (b) is that the student environment for learning is also radically divided by different epistemological commitments. Those who are, in principle, foundationalists are also rote-mode learners; those who are constructionists are "pattern or meaningful learners." I should also mention that we have some evidence, albeit limited in terms of numbers, that a gender difference exists: males are rote-mode learners and foundationalists; females are meaningful learners and constructionists.[6a] We need much more research here. Are Evaluation Researchers also so divided? What epistemological commitments do they have? What ideas about human learning, especially that of the handicapped, do they have? Does this difference, as it exists now, alter power relations? Does it influence what research we do? What grant monies are given? What are the social effects of different epistemological commitments?

An old maxim I learned long ago about evaluation is again relevant. "The politics that puts evaluation on the agenda is not itself on the agenda." In other words, the relation between power and evaluation knowledge is kept secret. Now, "secret" may be too strong a word; but the fact of the lack of clear, explicit relation between power and knowledge seems worth pondering. What research can we do to answer the question: "Why is it we have the knowledge we now do have?

Epistemology

Epistemology, the study of the structure of knowledge, is the broad point of view I take to list topics of needed research in evaluation studies of Transition Projects. Within Epistemology today two very different viewpoints are presented. One view is sometimes called Foundationalist, and it is argued that the right approach to knowledge is to find the certain (doubt-free) foundation of knowledge. For Empiricists, this foundation consists in sense-data, in observation, in a set of objectively neutral relations. Facts are discovered; a data-base is established; what is indeterminate is rendered determinate by research. Objective knowledge is highly desired. Objective knowledge, because it renders events into measurements (numbers, calculations), provides the only secure basis for social agreement. If, for example, you and I can agree that a "seven" is larger than a "five," then we have an objective basis for agreeing that a "seven" is better than a "five." In the context of poker, for example, and excepting designations of wild cards, sevens are higher than fives.

[6a] See Carol Gilligan, 1982 and Harding and Hintikka, 1983.

Constructionist views of knowledge are different. As the mention of the game "poker" suggests, what counts is what is, in a matter of prior and arbitrary agreement. In some contexts and conditions a five is more valuable than a seven. Constructionist views claims that knowledge, just as any result from evaluation research, is a human construction. Perhaps it would be a bit more sophisticated to suggest that the meaning of knowledge and value claims is a human construction. As we live in a semantic epoch (rather than the 17th century's Epistemic epoch, or the Greek Ontic Epoch[7]), we must pay attention to meaning. Philosophical differences turn on the idea that the truth admits of different valid formulations. Since truth is the top value of epistemology, different ideas of truth give different structure to research methodology. All of this evaluation into metaphilosophy can be confusing. Complexity overrides simplicity. Again, those who are frightened of certainty, the quick fix, the easy answer, the handy-dandy hologram, are our allies. They lift the rock on the by-way to see the ecology of the underside. Needed research in these areas needs researchers who are frightened, determined, calculating; persons-who-think- otherwise, persons who believe that getting smart about any human event that matters is not driven by the merely political, or methodological, or epistemological. While getting smart requires a lot of discipline, it is no longer a matter simply for the "Disciplines." It is a matter for human distinction.

---

[7] The Architectonics of Meaning, 1985.

Behaviorism. By way of historical explanation, we know the past 50 years of research in these fields has been dominated by Behaviorism and Positivism. Behaviorism, in principle, declined to look into the little "black box," the human mind. Behaviorists did not conceive that human thinking and feeling could be made objective, reliable, researchable. I believe we know differently today. I know concepts are what most people most of the time think with, and concept maps are a technic to externalize such thinking. I believe you can trust people in interview situations (both videotape and audiotape interviews). We have some fascinating videotapes of students in laboratory situations in which they comment freely on their observations of their own actions in the labs. They refreshingly indicate when they lied, faked it, or otherwise made devious moves from the moves which indicated their felt-significances, the times when feelings and meaning came together. So we know we can do better than Behaviorism/ Positivism. It may well be that the psychology of Behaviorism and its epistemology of Positivism have prevented us from "finding the problem." These now discredited views themselves may have "handicapped" those interested in helping the handicapped.

A research recommendation, then, is to study learning in the context of transition projects. The role concepts play is extremely critical. Every person who has a language has a rich store of concepts. And, intellectual power. Learning a language in the first place is perhaps the most difficult learning any person does. And it is done by the age of three or so. Learning a language, and using it to anticipate a future, is what sets apart human animals from all the others. Moreover, human beings learn many languages--just think of the languages of social memory, of music, of drama, of mathematics, of craft, of gestural-sign language, of clothing-dress-adornment (this cross stands for Christianity), of media- advertisements ("This Bud's for You"), and so on. The language of courage- in-the-presence of handicaps is a very special language. It is very much like the language of poverty in the presence of prosperity. Or the language of battlefield wars in the absence of peace. Evaluators need to find out more about the language of the handicapped operating in the presence of the not handicapped.

Organizing a Plan of Research.

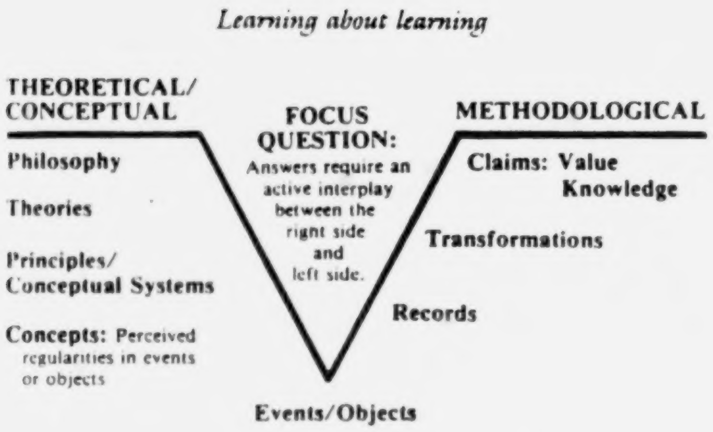
Organizing a plan of research is a creative process. As a graduate student in Stanford's Ernest Hilgard's seminar in psychological research, I learned what I now call Hilgard's maxim: whatever amount of time you have to do research, spend the first 75% thinking, the rest doing it. It is a difficult maxim to follow. Today, as a professor guiding research plans of graduate students, I begin with heuristics. The first, the Q-5 technic, requires the student to try to answer five questions. I recommend to those in this Evaluation Research effort to try to answer the following questions which follow:

A. The Q-5 Technic

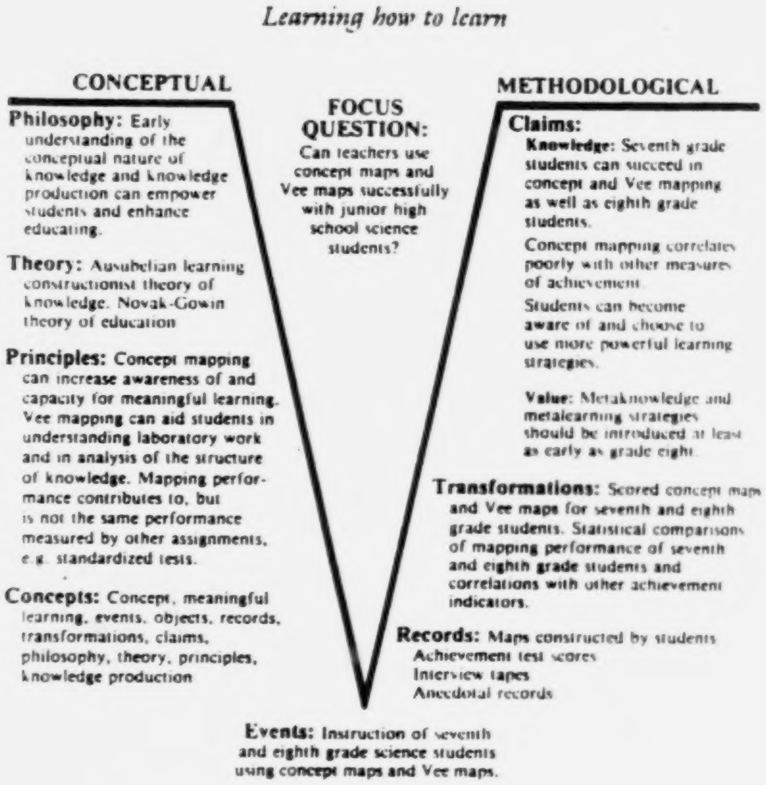
1. What is your focus question? or questions? What event/ object does the focus question focus on? Describe the context in which questions have meaning.
2. What concepts do you use to frame the focus question? Can you make a concept map of your key concepts. Are there any principles to give order and coherence to these concepts?
3. What methods, procedures, technics will you use to try to answer your questions? What events, specifically, will you make records of?
4. What would count as an answer to your question(s)?
5. What value (justification) does your research have?

Q-5 is only a beginning. As my students and I get a grip on the Q-5, we begin next to construct Vee diagrams. Often several Vee's are created. Some are too simple, some too complex. Remember, however, to keep these Vees, because as research progresses, hindsight may see that any early, simple Vee is the best.

The Vee Diagram Heuristic. The Vee diagram is a way to connect the dozen or so elements (epistemic elements) required in any complete research project. I present here a simplified Vee, and a more complicated one.


Learning about learning
THEORETICAL/ CONCEPTUAL
FOCUS QUESTION:
Answers require an active interplay between the right side and left side.
METHODOLOGICAL
Philosophy
Theories
Principles/ Conceptual Systems
Concepts: Perceived regularities in events or objects
Claims: Value Knowledge
Transformations
Records
Events/Objects

Vee One.[8]


Learning how to learn
CONCEPTUAL
FOCUS QUESTION:
Can teachers use concept maps and Vee maps successfully with junior high school science students?
METHODOLOGICAL
Philosophy: Early understanding of the conceptual nature of knowledge and knowledge production can empower students and enhance educating.
Theory: Ausubelian learning constructionist theory of knowledge. Novak-Gowin theory of education
Principles: Concept mapping can increase awareness of and capacity for meaningful learning. Vee mapping can aid students in understanding laboratory work and in analysis of the structure of knowledge. Mapping performance contributes to, but is not the same performance measured by other assignments, e.g. standardized tests.
Concepts: Concept, meaningful learning, events, objects, records, transformations, claims, philosophy, theory, principles, knowledge production
Claims:
Knowledge: Seventh grade students can succeed in concept and Vee mapping as well as eighth grade students.
Concept mapping correlates poorly with other measures of achievement.
Students can become aware of and choose to use more powerful learning strategies.
Value: Metaknowledge and metalearning strategies should be introduced at least as early as grade eight.
Transformations: Scored concept maps and Vee maps for seventh and eighth grade students. Statistical comparisons of mapping performance of seventh and eighth grade students and correlations with other achievement indicators.
Records: Maps constructed by students
Achievement test scores
Interview tapes
Anecdotal records
Events: Instruction of seventh and eighth grade science students using concept maps and Vee maps.

Vee Two.

---

[8] For a clear exposition of applications of concept mapping, Vee diagramming, and interviewing, see Novak and Gowin, 1984.

BEST COPY AVAILABLE

Evaluators studying their own practice need to study the utility of preparing their own Vees. They should compare Vees with other people. This moment is one of sharing meaningful thinking. It can be a creative moment. Some of my students have learned to bring a tape recorder to our sessions. Often we researchers do not recognize when we are being creative at just the moment it happens. Having a tape to record events also seems to free up creative thinking because you do not spend energy trying to remember what was said. Next, making a concept map (or maps) of audiotapes condenses a lot of talk to a few, good concepts. Again, be sure to keep these records to return to later when new difficulties turn up.

Vee diagrams also serve another value -- they show both what is well-grounded and what is missing. Efficiency of research is facilitated as you see (a) what is not a worry and (b) what is missing and, therefore, needs immediate work. These tasks are unfamiliar to many researchers. The only way to make sense of them is for researchers to use them; they are, after all, heuristics--a crutch, a help, a way to push understanding along.[9]

Specific Items of Research Interest.

Concept mapping is a very flexible and useful way of externalizing what is in one's mind. As such, this too, along with Vee diagramming, is handy for researchers at all aspects of research. Coupled with the audiotaped interview, it is a good way to understand clients. I would especially like to train some evaluators to interview handicapped clients in what I call "self-constructed dilemmas." That is, I would like clients to talk to evaluators about what troubles them, their discomforts/comforts,

---

[9] See Educating (Gowin, 1981).

the range of concepts they bring to bear upon their learning and working and living. We expect each individual's interview to result in concept maps (prepared by the evaluator) that are unique, idiosyncratic. But evaluator-researchers can read, say, a dozen "client-maps" and get something like a valid generalization about local knowledge.[10] That is, they can begin to read off these maps the recurring concepts, the recurring feelings-and-concerns of people in a specific context. I should write in passing that I think local knowledge is most valid, that it seldom generalizes to "national" knowledge per se, that actions taken in one context turn out to be valuable
but the "same" actions taken in another context turn out to be disturbing.[11] In other words, artifactual regularities can be formulated, but their specific meaning varies with context of use.

Services.

Evaluators evaluate services. I am troubled by this word, services. To be able to serve others is a value. But, what is the quality of the service? Is it good? Poor? Expedient? Poorly done on purpose? Is it like a funeral service, a social need ritualized and costly? I believe the key to quality of service is found in the idea of sharing meaning. So, I would ask evaluation researchers and project directors to scrutinize the meanings that are shared between the "served" and the "serving." This is one way to peer into the quality of the service.

---

[10] Local knowledge is discussed in a collection of pieces of interpretative anthropology by Clifford Geertz, 1983.

[11] See Michael Ritter & Barbara Tizard, 1979.

Dishonesty and Responsibility. Some of our studies, which stretch from junior high school through college levels, show a surprising and shocking concern: *rampant dishonesty about learning*.[12] We did not set out to study "dishonesty" and "irresponsibility." We were not looking for evils, for fault or blame. We were looking for ways people learn about learning. What we found is that our students in schools and colleges and universities have very limited ideas about learning. They readily admit it in interviews. They find most offerings of schooling events to be: boring, demeaning, insulting, not-worth-their-time; in a word, virtually meaningless. Yet the pressure to find "success" in school and life is so powerful, they find ways around the boredom. They fake it. They divert sixties is again relevant: our youth are growing up absurd. serious discussion into simplistic talk. They defy their own curiosity. They shut off new learning, especially if it appears difficult (and to require sustained effort). They have not the character for learning. Sadly, the schools did not teach them to learn about learning.

We are not talking here about moral idiots. We are talking about plain vanilla American young people. They desire to do well. They want the good life so well-displayed in consumer advertisements. They do not mind working; they expect to work. But when the work turns out to be insulting to their minds, they find ways to avoid it. A phrase out of the Unfortunately, the absurdities of the late 80's and the new 90's are not yet well-named. We could use some brilliant social analysis of these problems. Erving Goffman--where are you when we need you again?

---

[12] Cf. pp. 9, 10 and 11 in *Learning How to Learn* for a discussion of honesty and responsibility, and the ethics of teaching and learning.)

Handicapped youth are different. Perhaps, in an unusual way, they may recall us to our senses. Having what is labeled in social contexts as a "handicap" also labels these young people as different. They know they are different. As they mature to recognize their plight, they begin to develop character; they have to be honest about themselves. If I am blind, or deaf, or crippled, I find the spotlight focused on my disability; I am not normal. And..wonder of wonders...these "deviants" decide they cannot fake it. "It is true that I limp," "I need a crutch," "Coke-bottle eye-glasses distort my face," "A wheelchair is where I stand." And so on. Perhaps some of these young people have also been given symbolic wounds such as "Learning disabled," "Developmentally disadvantaged," and so on. I grew up on the folk-saying, "If you give a dog a bad name, then you can hang him." Vicious. Punish, to thereby discipline. These social devices are certainly not the wine and roses of romantic views of learning, of success, of knowing that money is life's report card. Those who have it easy, are easy; those who are tested in crucibles of adversity are thereby tempered; the unexpected consequence is that often these well-tempered souls are gracious, and value-giving. It may turn out to be that the most mature learners in schools and colleges are those a "success-means-normal" society labels "handicapped."

Schooling Fails the Employer's Test. A recent study of what employers think about high school graduates is not encouraging.[13] The employers have two complaints: graduates do not know how to be responsible for their work; they also do not get along well with other workers. Employers are not scholars or teachers primarily. They do not know much about learning or how to create a work situation such that a new employee learns. Yet, they find themselves thrust in such roles because the schooling system did not teach its students to learn about learning. Forget the chemistry, the English, the social studies. This typically out-of-date information is useless to those in the fast-moving world of work. Even high school computer studies are out-of-date in a very short time. Preparing students for work is extremely problematic today. Specific information and specialized skills are not an adequate preparation. The schooling system should forget such expedient goals. What the schools can do that is sorely needed is to focus on how we learn. If students get to learn about learning, to get knowledge about knowledge, then they have concepts, principles and approaches that are apt in many, many specific situations. Evaluate it.

---

[13] See "The School's Role in the Transition From Education to Work.

Transition Or Transformation.

These next comments lead to questions about the meaning of "transition." Is transition to be understood as a simple moving from one place to another? (As, for example, "The household goods are in transition from an apartment in Champaign to a house in Urbana.") Or, is some transformation in the quality of events expected, as in the judgment that a house in Urbana is a better place to live than an apartment in Champaign? Philip Jackson's 1986 talk at AERA contrasted the mimetic with the transformation. Both qualities are present in educative events. The mimetic is dominantly the occasion in schools, and it effects the poor more than the well-to-do. The transformative truly changes the meaning of human experience, and schools which achieve transformative qualities are rare. But this distinction between transition and transformation should lead us to ask questions about the quality of education transition projects achieve.

Schooling failed these handicapped. And if working is just like schooling, then working will also fail them. So what could happen in transition projects is merely to make a mechanical move from one social space to another, in which the quality of life is still the same, defeating. A good research question, then, is to study (a) the conditions that make schooling fail these people with handicaps, and (b) the conditions of working which are the same guides to failure.

What differences are there between "holding a job" (labor), and "creating wealth" (having a work, a calling as in a "vocation"). Labor is a task that routinely has to be done over and over again; washing the dishes after use is a labor. Shoveling snow does not add up to much, either, but it is a necessary labor. Working that creates wealth is different. Working, like educating but unlike schooling, changes the meaning and value of human experience.

Concluding. I realize that I am optimist, an idealist, an educator, words describing many of my colleagues. As educators we desire to increase intelligence, idealism, and wisdom, if possible. Such aims may not be achievable in the context of transition projects. The projects may fail, along with their clients. Sadly, it is the case that not all problems have solutions.

Footnotes

1. For a different view of events, as seen by a contemporary philosopher (who died recently), see R. M. Martin, 1978.

2. See Ulric Neisser, 1982. Neisser writes "The orthodox psychology of memory has very little to show for a hundred years of effort, perhaps because it has always avoided the interesting issues. Just as the naturalistic, ethnological study of animal behavior has proved to be more rewarding than traditional research on "learning," so a naturalistic study of memory may be more productive than its laboratory counterpart" (p. 3).

3. See also "Ecological Niche Picking," by Michael Cole, Lois Hood, and Ray McDermott (in Neisser's edited volume, p. 366 and passim). It should be obvious that "ecological validity" is another name for context and culture. Add to this reading that of Margaret Donaldson, 1978, who put "meaning" into a dozen or more Piagetian-type experiments and shows that if meaning is present, then even very young children are "formal operators." In other words, Donaldson destroys Piagetian certainty of stages and the assumptions of Piaget's genetic epistemology that "stages" are based on.

4. Cf. Michael Coles, 1985.

5. An incisive introduction to Foucault is Mark Philip's piece in the Quentin Skinner edited volume, 1985. The annotated bibliography is a good guide to further Foucault reading. I also recommend the other essays on social philosophers, especially the introduction by Skinner himself.

6. The two Richards, Rorty, 1980 and Bernstein, 1983, argue against Foundationlism (as well as relativism).

6a. That it seems to have a strong gender difference suggests a re-reading of Carol Gilligan's In A Different Voice. Gender-sensitive philosophers should read the Harding and Hintikka volume, Discovering Reality.

7. The Architectonics of Meaning, 1985.

8. For a clear exposition of applications of concept mapping, Vee diagramming, and interviewing, see Learning How to Learn (Joseph Novak and D. Bob Gowin, Cambridge University Press, 1984).

9. See Educating (Gowin, 1981.) Chapter 4, for definitions of the event (constructionist) epistemology in the context of educating. The Vee diagram is clearly presented here.

10. Local knowledge is discussed in a collection of pieces of interpretative anthropology by Clifford Geertz, 1983. It leads off with a most significant essay about academic disciplines, "Blurred Genres." It is through how we understand meaning (interpretation, signification) that we come to grasp a meaning not our own.

11. See 15,000 Hours (Michael Ritter & Barbara Tizard, 1979 is a report of three or four years of study of schooling in England. The main conclusion is that events-actions which are judged to be educative in one school are not so evaluated in another school. Meaning seems context-specific.

12. Cf. pp. 9, 10 and 11 in Learning How to Learn for a discussion of honesty and responsibility, and the ethics of teaching and learning.)

13. "The School's Role in the Transition From Education to Work," 4,078 employers who say high school experiences are "surprisingly unimportant."

References

Bernstein, R. (1983). Beyond objectivism and relativism. Philadelphia: University of Pennsylvania Press.

Coles, M. (1985). Intelligence as a cultural achievement. In E. Eisner (Ed.), Learning and teaching the ways of knowing. NSSE Yearbook. Chicago: University of Chicago Press.

Donaldson, M. (1978). Children's minds. New York: W.W. Norton.

Geertz, C. (1983). Local knowledge: Further essays in interpretive anthropology. New York: Basic Books.

Gilligan, C. (1982). In a different voice. Cambridge, MA: Harvard University Press.

Gowin, D. B. (1981). Educating. Ithaca, NY: Cornell, University Press.

Harding, S., & Hintekka, M. B. (Ed.). (1983) Discovering Reality. Hingham, MA: Kluwer, Boston.

Martin, R. M. (1978). Events: Reference and logical form. Washington DC: The Catholic University of America Press.

Neisser, U. (1982). Memory observed, remembering in natural contexts. San Francisco, CA: W.H. Freeman.

Novak, J., & Gowin, B. (1984). Learning how to learn. Cambridge, MA: Cambridge University Press.

Philip, M. (1985). Michael Foucault. In Q. Skinner (Ed.) The return of the grand theory in the human sciences (pp. 65-82). Cambridge, MA: Cambridge University Press.

Ritter, M., & Tizard, B. (1979). 15,000 hours. London: University of London, Institute of Education.

Rorty, R. (1980). *Philosophy and the minnoi of nature*. Princeton, NJ: Prineton University Press.

Watson, W. (1985). *The architectonics of meaning*. Albany: State University of New York Press.

The school's role in the transition from education to work. Research Dissemination Office, Baltimore, MD: John Hopkins University, Center for Social Organization of Schools.

NEEDED RESEARCH ON OPTIMIZING THE STRUCTURE OF EVALUATION DESIGNS: THE ISSUE OF FLEXIBILITY IN EVALUATION WORK[1]

NICK L. SMITH
Syracuse University

An evaluation design serves many purposes. As a protocol for an applied inquiry study, it specifies the actions required to obtain valid and reliable information. As a management plan, it can serve as a tool to guide the implementation and completion of the study. As a statement of professional obligation, it can help to assure clients that useful and timely information is provided in accordance with prior agreements. Indeed, one can construe the primary purpose of an evaluation design as a device for insuring that certain qualities valued by various parties are preserved throughout the evaluation effort (Smith, 1982a).

An evaluation design is a customized statement of intent to accomplish the purpose of the evaluation study, within a specific context, and in light of the nature of the program being evaluated. As the purpose of the study, its context, or program characteristics change, so might we anticipate that the design of the evaluation will have to change in order to optimize the design and to insure that the desired qualities of the study (e.g., validity, utility, timeliness, etc.) are preserved.[2] Further, if the program operates within a dynamic context or is itself developmental and still changing, it may be difficult to prespecify what an optimal design would be. (See, for example, Tripodi, Fellin, and Epstein, 1978, for a discussion of the relationships between different evaluation strategies and a program's stage of development.) To what extent can or should an evaluation design be adaptive to these conditions?

In short, optimizing the structure of a design requires that it be sensitive to a variety of environmental factors (e.g., context, purpose, resources, client needs, program characteristics, etc.) which may change during or after the design formation process. It is important, therefore, to consider ways in which reasoned changes can also be made to the evaluation design. In what ways can or should an evaluation design be flexible to accommodate a changing environment?

Accumulated experience in evaluation utilization suggests that a certain degree of flexibility is beneficial in maximizing services to clients. Also, more open, emergent designs have been advocated by those studying "naturalistic" approaches to evaluation. But little formal research has been conducted on the issue of design flexibility, its variations, implementation, and effects. Some practitioners are able to develop and implement designs that are methodologically strong, useful to clients, and also resilient to changing environmental conditions. How do they do that? What can the evaluation profession learn from current practice that could be used to build empirically-based recommendations for improving the flexibility of evaluation designs? The purpose of this chapter is to elaborate on the nature of flexibility in evaluation work and to propose ways in which the study of evaluation flexibility would improve the evaluation of school to work transition programs for handicapped youth.

Flexibility and The Evaluation Of Transition Programs

Design flexibility is an important methodological issue in general, but it is a particularly relevant concern in the evaluation of the current transition programs for handicapped youth.

The evaluation work within these programs is highly vulnerable to environmental changes. There is great variety in the types of these secondary/transition programs in terms of student or client characteristics, operational settings, geographic spans of services, resources, and periods of operation (Heal, Haney, Bamman, and Phelps). This environmental variety means that a common evaluation design is not possible, as evidenced in the various designs proposed in the initial grant applications. Further, a common, adequate response will not be likely for any shared environmental change, such as changes in Office of Special Education and Rehabilitative Services (OSERS) regulations or requirements. Each program will have to deal independently with any necessary evaluation changes, whether caused by events shared or unique to each program.

Further, these programs have allocated few resources to evaluation, with approximately a quarter of them proposing to hire an outside evaluator at less than ten percent time. If conditions require alterations in the existing evaluation designs, these programs will not have the needed evaluation expertise at hand and will have a difficult time effecting suitable design changes. Only experienced evaluators with considerable involvement in the programs would have been able to initially design flexible studies and would now be able to make the unanticipated changes as the need arises.

One of the purposes of the Secondary/Transition Intervention Effectiveness Institute at the University of Illinois, Urbana-Champaign, is to provide technical assistance on evaluation methods to these programs. Without more knowledge about the nature of flexibility in evaluation work and the principles which determine what design alterations are possible or desirable, however, the Institute staff can only assist the programs on a case-by-case, ad hoc basis. If it is likely that the programs will have to make alterations in their evaluation work, then some conceptual and empirical research on evaluation flexibility is necessary in order for the Institute to provide adequate technical assistance to the diversity of transition programs.

And changes will almost assuredly need to be made in the programs evaluation work. It is typical in large collections of federally-funded local projects that changes in federal resources or policies necessitate changes in local level evaluation efforts. This has been the experience, for example, with Title I/Chapter 1, Follow Through (cf. House, 1979), and Categorical/Block grant programs (cf. Gray, Caulley, and Smith, 1982). Indeed, questions of reauthorization and level of future funding have already changed the expectations OSERS has for the evaluation results to be produced by the local programs. Further federally-initiated changes can reasonably be expected.

Even without OSERS occasioned changes, it is probable that the local programs will themselves feel the need to alter their evaluation designs. Some programs undoubtedly gave little early thought to their evaluation plans, as evidenced by grant application descriptions as short as 5 sentences. Unanticipated problems, such as the inability to locate suitable assessment instruments, unreliable rating scales, and unavailable employment statistics will necessitate design alterations. Obtaining suitable evaluation help, especially at less than ten percent time, may prove difficult; program directors who listed themselves as primarily responsible for the evaluation may find they have too little time to devote to it; and some proposed methods, such as cost-benefit analysis, may prove to be much more difficult to implement and interpret than originally anticipated (see, for example, Denton and Smith, 1985).

Events at both the federal and local levels are, therefore, likely to require changes in the evaluation work at the program level. Insufficient resources and expertise at the local level will result in an increased need for technical assistance from the Institute, but generalized solutions will be possible only if the Institute staff has a better understanding of the nature of possible changes and of their probable impact on subsequent evaluation work. Studies of design flexibility are therefore needed in order for the Institute to adequately fulfill its technical assistance role.

There are two additional reasons why design flexibility is a desirable research topic for improving the evaluation of transition programs: its long-range potential for improving future intervention programs for the handicapped, and the opportunity it affords for integrated studies with other research topics being proposed by the Institute.

The Institute has attempted to identify generic evaluation issues of broad applicability whose resolutions promise important contributions to evaluation generally, as well as enable significant improvements to be made in the evaluation of the current transition programs. To focus exclusively on the immediate evaluation problems of the transition programs is to lose the opportunity for a larger, more powerful understanding that may not only assist the present programs, but also contribute to the improved evaluation of future intervention programs for the handicapped, and to better evaluation methods in general. Given the uncertain future of the transition programs in the Congressional reauthorization process, it would be prudent to conduct research on evaluation issues which have the potential of improving both the evaluation of the current transition programs as well as program evaluation more generally. Design flexibility is such an issue. Research on evaluation flexibility will insure that a contribution is made to the improvement of programs for the handicapped, regardless of whatever Congressional decision is reached.

Finally, in researching design flexibility, the Institute would have the opportunity to conduct integrated studies to examine the interrelationships of several of the other issues being proposed for research; for example, context, values, and standards. One of the key questions concerning design flexibility is how changing or competing contextual factors can be accommodated within an existing design. More specifically, how can the transition programs provide information usable within the federal context, based on procedures originally established to satisfy the needs of the local context? To some extent, it is the differing values of the participants within these two contexts which give rise to competing demands on transition program evaluations. It is important, therefore, to understand better how values influence evaluation methods (e.g., House, 1979), and the ways in which evaluation procedures can be constructed to accommodate competing values (e.g., Smith, 1982). Some criteria or standards must be invoked in order to assess the quality of the evaluation changes which result from these accommodations to competing contextual values. Therefore, the appropriateness of certain standards in judging the adequacy of design changes must be explored. Context, values, and standards in evaluation are all worthy of research in their own right, but the study of design flexibility permits an examination of their interrelationships in improving the evaluation of transition programs for the handicapped.

Having discussed the need for research on flexibility, the remainder of this chapter elaborates on several types of evaluation flexibility, considers factors that influence flexibility, and suggests the kinds of research studies most appropriate at this time to understand evaluation flexibility better.

Flexibility in Evaluation Work

To develop a research agenda on evaluation flexibility, it is first necessary to clarify what is meant by flexibility in evaluation work. In the following sections, several terms are defined and a framework for thinking about flexibility in evaluation is provided. While the following are only preliminary thoughts and more conceptual development is needed, they do provide a sufficient basis for initial field studies on evaluation flexibility. Such studies would provide the knowledge necessary for a more sophisticated view of design flexibility.

The Random House Dictionary (unabridged, 1967) defines "flexible" as "susceptible of modification or adaptation." "Flexibility" is defined in this chapter as a neutral, descriptive characteristic of evaluation work; that is, it is neither inherently desirable nor undesirable for an evaluation to be flexible. There may be purposes, contexts, and procedures under which high flexibility is desirable, and others under which low flexibility is better. Evaluation flexibility is, therefore, not defined as an absolute good, something always to be desired in evaluation such as fairness of relationships or validity of information, but as a conditional good, something desirable or beneficial under some conditions but not others. A key question under the research agenda, then, might be "Under what purposes, contexts, and procedures in the evaluation of these transition programs is flexibility desirable or undesirable?"

Since flexibility is a continuous, not a dichotomous, attribute, evaluation work can be viewed as varying from fluid to formable to firm to fixed, requiring one to think in terms of degree of flexibility. It would then be more precise to ask "What level of flexibility is best under given purposes, contexts, and procedures in evaluating transition programs?"

There are two broad categories to consider in thinking about flexible evaluations, those situations in which several, smaller, discrete evaluation designs are employed in evaluating a program (multiple-design flexibility), and the more common case in which a larger, single design comprises the evaluation (single-design flexibility). Each category is considered below.

Multiple-Design Flexibility

One way in which evaluation work can be made flexible is by conducting a series of small studies, each with its own separate design, and each constructed on the basis of the results of the previous studies.

This approach is sometimes employed in the single-subject designs used to evaluate clinical therapies (cf., Kratochwill, 1978, Bloom and Fischer, 1982). Although single-subject studies are often single designs constructed for causal hypothesis testing, they may also consist of multiple-design studies used to provide evaluative feedback to improve clinical practice. For example, a sequence of single-subject designs might be employed to (a) institute a therapeutic treatment and test its impact on behavioral change, (b) introduce a second treatment to determine if it has an additive or inhibitory effect on the behavioral change, and (c) then assess the extent to which the desired behaviors extinguish upon withdrawal of the treatments. This design strategy is different from initially constructing a single, stagewise design in that each step is designed only after knowing the results of the previous steps. For example, if the first treatment is ineffective, a second treatment would not be added to it, but might be substituted for it instead. And, of course, an attempt to see if the desired behaviors might extinguish would not be made unless one or more of the previous treatments had been effective in establishing the behaviors.

Multiple-design studies are not particularly prevalent. Perhaps evaluators simply find it too difficult, time consuming, and costly to implement single-design studies to consider conducting a series of studies, even if they are smaller and more targeted. Variations of the multiple-design strategy are occasionally used, however, as when a long term curriculum evaluation project conducts a needs assessment, followed by a materials review by experts, followed by in-house formative testing, and finishing with a summative field evaluation.

The application of the multiple-design strategy to the evaluation of programs for the handicapped has been treated in some depth by Smith (in press). He describes the rationale for employing a series of small studies in evaluating preschool programs for handicapped children (the other population most focused on in P.L. 98-199, along with the youth of the transition programs). Several examples of how a given study can arise as a result of a previous investigation are provided there which illustrate the flexibility of the approach and its responsiveness in meeting changing contextual factors and client needs. Because of this previous treatment, the multiple-design strategy will not be discussed further in this chapter.

Single-Design Flexibility

In what ways might a study employing a single evaluation design be flexible? There are three aspects of the evaluation process for which flexibility might mean different things: design formation, design structure, and design accommodation. Each of these topics is treated, in turn, below.[3]

Flexible Design Formation. "How much flexibility is desirable when initially creating an evaluation design?" Should the design be determined very early in the evaluation process; for example, even dictated out right as is done in some mandated studies, requests for proposals, and state legislated evaluations? "What are the costs and benefits of a preordinant design approach?

"Which elements of the design (clarification of client relationship, identification of audiences, crystallization of evaluation issues and questions, selection of information sources, development of data collection instruments, selection of analysis procedures, determination of reporting formats) can be decided on in a more fluid manner, and which elements should soon be fixed?" For example, does the client always have to be identified early in the study, or is it ever better to clarify the major evaluation issues and identify the range of audiences before selecting the client? Is it ever good practice to first establish a data collection, analysis, and reporting system, in a preordinant fashion, and then let the specific audiences and concrete evaluation questions emerge over time? (Indeed, isn't this the model followed by state departments of education and large school districts when they rely on standardized management information systems to respond to requests for evaluation information?) How soon do the evaluation design elements need to be decided and in what order?

Wnen should one delay in deciding on the design elements for as long as possible? Perhaps particular knowledge or a certain condition is required before one can determine an evaluation design, and an emergent strategy is employed until that knowledge or condition is obtained. If so, what is the knowledge or condition that is needed? In general, "What are the costs and benefits of an emergent design approach? In particular, what kinds of things should a transition program know before finalizing its evaluation design?

Can evaluation work be simultaneously both preordinant and emergent, as when some design elements are fixed early and others decided as seem warranted? Some evaluation contractors who do repeated work for a group of clients seem to design individual studies in a preordinant style, but allow their relationships with clients, their specialization on specific evaluation problems, and their methodological expertise, to emerge over time. Generally speaking, "Under what purposes, contexts, and procedures is it better to employ a preordinant approach to determining an evaluation design, and when is it better to follow an emergent approach?"

Flexible Design Structure. One can construct evaluation designs that are more or less flexible; that are designed to be predominantly fluid and changeable as conditions warrant, or are designed to be fixed and relatively unchangeable over the period of the study. "Under what purposes, contexts, and procedures should evaluation designs be more or less flexible?" For which transition programs should the evaluation designs be fixed, or should they all be fluid to accommodate the changes likely to be required?

Note the difference between flexible design structure and flexible design formation. One can employ an emergent approach to forming a design which results in the decision that a fixed design structure is best in this case. Or, one could decide at the outset (preordinant) that a fluid design which will allow the evaluators to respond to the constantly changing information needs of the program is required. Probably the most commonly occurring combination is the preordinant decision that a fixed design is best. In the final case, where an emergent process is used to select a fluid design structure, it may be difficult to identify just when the actual design decisions were made.[4,5]

Since an evaluation design is defined here as a discrete plan of action for conducting an evaluation, one can ask "Which design elements can remain fluid throughout the evaluation study and which need to be more firm or even fixed?" Would one ever plan to change the client or primary audience as the evaluation proceeded? (One might do so, for example, in a long term evaluation with a planned formative to summative shift.) What are the costs and benefits of allowing the data collection and reporting strategies to be periodically renegotiated by the client? Under what conditions will continuing to "refine" the evaluation questions or issues result in a more targeted study, and when will it endanger the study's continuity and cohesiveness?

Flexible Design Accommodation. The third aspect of evaluation in which to consider flexibility is the accommodation of the evaluation design to unforeseen events or conditions. "What is the flexibility of different evaluation approaches in accommodating to unforeseen changes?" The difference between this form of flexibility and that of the previous section is the difference between planned and unplanned changes. Presumably, if an evaluator can anticipate events that are likely to alter the design, he or she would attempt to control them if using a fixed design, or change to a more fluid design to accommodate them. Since the need for unplanned changes often arises, however, we must ask about the flexibility of a design to incorporate such changes without seriously jeopardizing the design's integrity.

The form of flexibility, then, refers to a design's robustness. What kinds of unforeseen conditions or events might arise, and how might they affect the design? Common occurrences which force design changes include dramatic reductions in funding, loss of key personnel, inability to gain access to essential information, mandated alterations in the primary purpose or scope of the study, etc. Are some designs more robust to these kinds of influences than others? Are there techniques for increasing a design's flexibility to deal with such events? To what extent should experienced evaluators be able to anticipate such events and construct more flexible studies to begin with? (For example, knowing from experience that funding cuts in federally-supported studies are common, one might modularize a previously integrated design in order to have salvageable components should the funds be reduced. If the budget cuts did occur, one would then be able to regroup more easily into a still serviceable design.)

Although the need for flexibility in making unplanned changes often arises in evaluation practice, little is known about how to cope with this recurring problem. House's (1979) metaevaluation of the Follow Through experience, for example, chronicles the considerable difficulties in making such changes. Further, it is not clear whether fluid or fixed designs are more robust in responding to unforeseen changes. Changes that require greater control and rigidity of structure may be as difficult for a fluid design to accommodate as changes that force a fixed design into a more fluid configuration. "How do we increase the robustness of evaluation designs?"

These are especially important considerations for the transition programs. What conditions are they currently responding to that may alter their evaluation designs? What changes are anticipated? What are the characteristics of the programs that are able to effectively adapt their designs? Are there general strategies that the transition programs can use to modify their designs to comply with new OSERS requests but still satisfy local information needs? How can the robustness of the transition programs' evaluations be strengthened?

The flexibility of evaluations can be considered from three different viewpoints: flexibility in the formation of a design (e.g. preordinant and emergent design strategies), flexibility in the structure of the design itself (fluid to formable to firm to fixed), and flexibility of the design in responding to unforeseen events (robustness). The general research question for all three types of flexibility is, "What are the relationships between purpose, context, clients, procedures, etc., and desirable levels of evaluation flexibility?"

Factors Influencing Design Flexibility

A number of factors may require that evaluation designs be flexible: for example; changes in study purpose, context, clients, resources, or requirements; a lack of particular knowledge needed to finalize a design; or the desire to base subsequent evaluation decisions on the results of prior efforts. What has not yet been considered, however, is that there may be limits to how flexible certain designs can be. What might influence an evaluation's ability to be flexible? Two factors are briefly considered here, the basic evaluation models or inquiry strategies employed in a study, and the nature of the program being evaluated.

Models and Inquiry Strategies. Since many evaluators are trained to use the common evaluation "models" (cf., Madaus, Scriven, and Stufflebeam, 1983) in designing studies, it is important to understand, "What is the relationship between the level of flexibility in a design and the model used to develop it?" Are some models biased in favor of greater or lesser flexibility in constructing a design? Is the Tylerian model (e.g. Smith and Tyler, 1942; Tyler, 1949; Tyler, 1983) relatively more fixed because of its procedural specificity, preordinant focus on objectives, and presumption of which audiences are primary, as compared with a more fluid Goal-Free model (e.g. Scriven, 1973; Scriven, 1974) which allows the evaluative focus and criteria to emerge from initial observations? And what are the costs and benefits of this greater flexibility -- is the Tylerian model more likely to provide conclusive evidence while Goal-Free studies provide only suggestive or preponderant evidence? Are there particular models that evaluators can employ to insure a given level of flexibility?

Similar questions can be asked about the inquiry strategies commonly used in evaluations. "Do certain inquiry strategies require particular kinds of 'fixedness' or 'fluidness' in the evaluation design?" For example, experimental designs require stability of treatment, surveys require commonality of language, single subject designs require elapsed time to establish base-rates, naturalistic observation requires freedom from artificial intervention, and open-ended interviewing requires following leads as they arise. Do designs which maximize internal validity by controlling for alternative explanations thereby jeopardize the study's robustness by limiting the kinds of subsequent changes permitted in the design?

Answers to questions such as these may help identify which designs are most robust and can be implemented in highly changeable situations, and which designs are more fragile and should only be used when changes are unlikely or can be controlled.

Program Characteristics. Perhaps some program attributes limit the levels or types of evaluation flexibility that are possible. "What is the relationship between the flexibility of the evaluation design and the nature of the program being evaluated?" Consider, for example, the possible combinations of a fluid or fixed evaluation design and the fluid or fixed aspects of the program being studied. In evaluating the placement of computers in junior high school libraries, one might study the change over time in peer tutoring behaviors (a fluid program aspect) by using a fixed photographic study with preset data collection procedures and times, or an open-ended photographic strategy using fluid data collection procedures that are modified as the study progresses. Alternatively, one might study student attitudes about the software available (a fixed program aspect) using a fixed strategy examining frequency of use and student rating forms, or a more fluid interview strategy that progresses to increasingly specific questions about the most popular software. Are there decision rules for deciding when a more fluid or a more fixed evaluation design is better, depending on how fluid or fixed the program aspect is?

A number of useful investigations might be conducted, then, to identify the relationships between the evaluation models and inquiry strategies used by the transition programs and their ability to respond to needed design changes. Also, are certain transition programs, by their very nature, less flexible and less able to accommodate changes in their evaluations? It is important to know what factors may necessitate changes and when changes are desirable, but it is also critical to understand the reasons why certain types of changes may, in fact, not be possible.

What procedures can be used to optimize the match between a transition program and its context, and suitable levels of evaluation flexibility? How does one determine what level of flexibility is desirable in a given case? What is the relationship between the "initial condition" of a program and the need for evaluation flexibility of all three types? (Perhaps evaluability assessment procedures could be modified to aid in addressing these issues.)

There appears to be a great deal not yet known about flexibility and evaluation design. Some of the most important issues and questions have been identified in the text above.

Determining Desired Flexibility

The following are some preliminary thoughts on the issue of how one might determine what an acceptable level of flexibility in an evaluation should be. A general solution is proposed here; specific solutions in concrete cases would require answers to the conceptual and empirical questions raised above.

Consider the following principle for determining how flexible an evaluation should be: "The desired level of flexibility in an evaluation is that which optimizes the integrity of the evaluation work." The integrity of an evaluation is defined to consist of three elements: its Utility, Propriety, and Accuracy. These elements are roughly equivalent to the major categories of the Joint Committee Standards for evaluation, with the feasibility standards subsumed under Utility (Joint Committee, 1981). Utility has to do with the usefulness of the service provided by the evaluation, such as elaborated on in the Standards. Propriety has to do with the legal and ethical aspects of the evaluation, including such considerations as House's (1980) concepts of fairness and justice in evaluation. Accuracy includes those topics covered by the Standards as well as truth and validity in evaluation as discussed by House (1980). Therefore, the desired level of flexibility in an evaluation is that which optimizes the evaluation's utility, propriety, and accuracy. To the extent that flexibility weakens these aspects of the evaluation, it is not desirable.

For each of the three types of evaluation flexibility discussed above, one can now ask, "To what extent would greater flexibility increase the study's utility, propriety, or accuracy?" Therefore, under what conditions does using an emergent process for designing an evaluation increase its utility to the client? When does a fluid design provide greater utility and when is a fixed design better? How does maximizing a study's robustness affect its ultimate utility to the client? Similar questions can be asked concerning propriety and accuracy.

Questions of consumer protection and evaluation malpractice can be raised under the issue of propriety. Are certain levels of flexibility required in order to protect the clients of the program or of the evaluation? Under what conditions might creating a design without sufficient flexibility or robustness be considered malpractice? Could failure to anticipate and respond flexibly to unanticipated events be grounds for malpractice?

What is the relationship between an evaluation's flexibility and its accuracy? Do emergent design strategies result in more relevant but less precise information? Is the degree of certainty of an evaluation's results greater for fixed than for fluid designs? (See, for example, Smith, 1981, Morris and Sales, 1982, and Smith, 1982b, for discussions of the relationships among program characteristics, evaluation designs, and the certainty of evaluation information.)

Deciding the appropriate level of flexibility for a given study requires a holistic assessment of the trade-offs of the various types of flexibility in terms of the evaluation's ultimate utility, propriety, and accuracy. Such a judgment would be similar to an architect's consideration of "architectural fit" in judging the adequacy of a building. Given the purpose of the structure, the site characteristics, and the materials available, is this building a functional and aesthetic solution to the architectural problem (cf. Ferguson, 1981)? Similarly, one might ask: given the purpose of the evaluation, the nature of the program and its context, and the resources available, are the levels of flexibility in the design process, the design structure, and the design robustness, the best possible solutions for optimizing the evaluation's utility, propriety, and accuracy?

In order to make this assessment, however, one would need answers to the conceptual and empirical questions raised above. The following section, therefore, briefly discusses the kinds of studies needed to answer those questions in the evaluation of transition programs.

Studies of Evaluation Flexibility

The studies described below are those needed to build a more informed view of the nature and importance of flexibility in the evaluation of transition programs, and to develop a more refined understanding of the role of flexibility in evaluation in general. Although a variety of studies are possible, conceptual studies and investigative studies of evaluation practice are most needed at this point, given the dearth of conceptual clarity and basic information on the topic.

Conceptual studies are needed in order to clarify, revise, and expand the concepts, definitions, and framework suggested in this chapter. Clear cases, counter cases, and borderline cases from evaluation practice should be identified to test the validity and relevance of this analysis. The logical, ethical, and methodological implications need to be elaborated further. Also, the logical relationships of evaluation flexibility to the evaluation models, common setting and contextual characteristics, and common inquiry methods must be worked out. This research will provide the necessary conceptual basis for conducting field studies of evaluation practice.

Investigative studies of evaluation practice by the transition programs are necessary in order to gather evidence of empirical relationships, discover possible explanatory connections, and to record the perceived meaning of events to the program participants. These field studies of naturally occurring evaluations should examine the kinds of difficulties that evaluators and clients are experiencing related to evaluation flexibility, the events that necessitate design change, the factors that inhibit or promote flexibility, and the means that are currently being used to deal with issues of flexibility, as well as the success of their application. Studies of program responses to specific events, such as new OSERS requests for information, will be particularly helpful in enabling the Institute to provide direct technical assistance in solving problems the transition programs are experiencing.

These studies of evaluation practice are needed in order to ground the conceptual work in actual practice, to gather descriptive information on the nature, level, and effects of flexibility in practice, to gather information on the relationships of evaluation purpose and program context to evaluation flexibility, and to identify effective strategies for selecting and implementing the appropriate levels of flexibility. The results of these investigative studies of practice would include information on the costs and benefits of various naturally occurring levels of evaluation flexibility, recommendations for optimizing flexibility, and procedures for determining and implementing the desired levels of flexibility.

The results of such conceptual and investigative studies would add considerably to the evaluation profession's understanding of ways to improve evaluation designs, would provide valuable information concerning the evaluation of future intervention programs for the handicapped, and would enable the Institute to provide more effective technical assistance to the existing transition programs.

Footnotes

[1] This work was supported in part by the Secondary/Transition Intervention Effectiveness Institute at the University of Illinois, Urbana-Champaign; its contents are solely the responsibility of the author, however, and no endorsements should be inferred.

[2] One might argue that optimizing a design is too stringent a requirement for many field evaluations, especially the low budget efforts conducted by some of the transition programs. The argument in this paper applies equally well, however, whether one is optimizing a design or only satisfying. In the latter case, greater changes in study purpose, context, or program would occur before alterations would be needed to insure that a satisfactory design was maintained.

[3] The assumption is made here that a design is a discrete, identifiable entity that can be formed, implemented, and changed. If there are other ways of conceiving of evaluation designs, they are not considered here.

[4] An extreme variation of this case is where a discrete, identifiable design is never selected, but is ever "emerging". Under the assumption made in this chapter of what a design is, however, this would not be considered a formal design, although it certainly would reflect high flexibility.

[5] For the sake of illustrations, flexibility is often dichotomized in this chapter (e.g. preordinant/emergent process, fluid/fixed structure), but this is for convenience only. Flexibility is a continuum, and studies of evaluation flexibility should address not issues of either/or, but of how much.

References

Bloom, M., & Fischer, J. (1982). Evaluation practice: Guidelines for the accountable professional. Englewood Cliffs, NJ: Prentice-Hall.

Denton, J.J., & Smith, N.L. (1985). Alternative teacher preparation programs: A cost-effectiveness comparison. Educational Evaluation and Policy Analysis, 7, (3), 197-205.

Ferguson, G. (1981). Architecture. In N.L. Smith (Ed.), Metaphors for evaluation: Sources of new methods. Beverly Hills, CA: Sage.

Gray, P.J., Caulley, D.N., & Smith, N.L. (1982). A study in contrasts: Effects of the Education Consolidation and Improvement Act of 1981 on SEA and LEA evaluation (ROEP Paper and Report Series, No. 79). Portland, OR: Northwest Regional Educational Laboratory.

Heal, L.W., Haney, J.I., Bamman, V., & Phelps, L.A. (1986). First annual monograph meta-analysis evaluation group. University of Illinois, Champaign-Urbana: The Secondary Transition Intervention Effectiveness Institute.

House, E.R. (1979). The objectivity, fairness, and justice of federal evaluation policy as reflected in the Follow Through evaluation. Educational Evaluation and Policy Analysis, 1, (1), 28-42.

House, E.R. (1980). Evaluating with validity. Beverly Hills, CA: Sage.

Joint Committee on Standards for Educational Evaluation (1981). The standards for evaluation of educational programs, projects, and materials. New York, NY: McGraw-Hill.

Kratochwill, T.R. (1978). Single subject research: Strategies for evaluating change. New York, NY: Academic Press.

Madaus, G.F., Scriven, M.S., & Stufflebeam, D.L. (Eds.). (1983). Evaluation models: Viewpoints on educational and human services evaluation. Boston, MA: Kluwer Nijhoff.

Morris, R.A. & Sales, B.D. (1982). Legal standards for evaluation research: A reply to Smith. Evaluation and Program Planning, 5, 217-221.

Scriven, M.S. (1973). Goal-free evaluation. In E.R. House (Ed.), School evaluation. Berkeley, CA: McCutchan.

Scriven, M.S. (1974). Pros and cons about goal-free evaluation. Evaluation Comment, 3, 1-4.

Smith, E.R., & Tyler, R.W. (1942). Appraising and recording student progress. New York, NY: Harper.

Smith, N.L. (1981). The certainty of judgments in health evaluation. Evaluation and Program Planning, 4, 273-278.

Smith, N.L. (1982a). Evaluation design as preserving valued qualities in evaluation studies. Studies in Educational Evaluation, 7, 229-237.

Smith, N.L. (1982b). Levels of evidence and degrees of certainty in evaluation: A response to Morris and Sales. Evaluation and Program Planning, 5, 313-315.

Smith, N.L. (in press). Evaluation alternatives for early intervention programs. In L. Bickman & D.K. Weatherford (Eds.), Evaluating early intervention programs for severely handicapped children and their families. Austin, TX: Pro-Ed Publishers.

Tripodi, T., Fellin, P., & Epstein, I. (1978). Differential social program evaluation. Itasca, IL: F.E. Peacock.

Tyler, R.W. (1949). Basic principles of curriculum and instruction. Chicago, IL: University of Chicago.

Tyler, R.W. (1983). A rationale for program evaluation. In G.F. Madaus, M.S. Scriven, & D.L. Stufflebeam (Eds.), Evaluation models: Viewpoints on educational and human services evaluation. Boston, MA: Kluwer Nijhoff.

SITUATIONAL CONTEXT AS INFLUENCE ON EVALUATION DESIGN AND USE

Robert E. Stake
University of Illinois

> ..to speak of an effect [of an intervention] is short-sighted. The circumstances surrounding the intervention are part of the cause.
>
> Lee J. Cronbach, 1982

The purpose of this paper is to identify needed research on situational context which might lead to the improvement of evaluation. In these several papers we aim eventually at improving the evaluation of programs facilitating transition of handicapped youth from school to work. Much of what is presented here is applicable to program evaluation generally.

It is the task of all research, and of course of program evaluation studies, to give added meaning to matters. An object evaluated is an object whose value has been clarified. But the clarification is not only of its value; there is clarification also of the nature of the object. In program evaluation we necessarily come to understand programs in new and different ways.

To understand an object is to understand both its common and unique features. There is an abundance of each. Our study may emphasize the common or the unique, the generalizable or the local, but in any successful study both the common and the unique will be clarified.

And to understand an object is to understand both its integral and its contextual features. Some of its internal parts will distinguish it from other members of its class, some will be common and will help identify the class to which it belongs. Likewise, some of the contextual features of the object will reveal its uniqueness whereas other background characteristics will help identify the class to which it belongs (Schutz, 1963).

Educational programs, like most objects worth study, have important contexts: temporal, physical, spatial, social, political, economic, etc. History, geography and each of the social and humanistic disciplines identifies context which enrich the meaning of education. An educational practice has its habitat, its milieu, its frame of reference, its zeitgeist--not one but many contexts. One task of program evaluation is to draw attention to the particular situation.

Many contextual features are well watched by transition project personnel: e.g. the availability of jobs, the support of the School Board, the competence of the van repairman. A program evaluator sometimes finds these features contributing to success or failure. Still, in a figure-ground metaphor the program is figure, the contexts are ground. For other studies any one contextual feature might become the focus of attention, but for program evaluation it remains ground.

We know contextual features can be important, but our knowledge is casual, intuitive. What we do not know well enough and what we should make further effort to study is the epistemological role of context in the meaningfulness and utility of evaluation findings. Qualitative methods of research give particular attention to context (Brandt, 1972; Schatzman and Strauss, 1973; Wolcott, 1976; and Erickson, 1986). The evaluation work of Parlett and Hamilton (1976), Patton (1981), Guba and Lincoln (1981), and my own emphasis on the use of case study in evaluation (Stake and Easley, 1978; and Stake, 1981) are step stones. We need to know how much our research designs should call for resources to be spent on the examination of context. We know too little about when contextual emphasis will be more important and when less. This paper will take up three aspects of context:

1. Context as causal influence on program success.
2. Context as influence on willingness to generalize findings.
3. Structural versus contextual approaches to evaluation design.

Finding Explanation Influenced by the Program's Contexts.

Educational research is usually a search for generalizations, even causal relationships, among independent and dependent variables. It seldom seeks the particular contextual element as main cause of an outcome. The instructional researcher usually tries to find an abiding influence on achievement attributable to instructional pattern or student input, not something incidental such as whether or not the intestinal flu on test day made a difference.

Program evaluation should be interested in both abiding and incidental determinants of outcome. The teacher's relationship with the vocational coordinator, the preemption of the gym by the wrestling team, and community vocalization of the "evils" of social promotion may be important ingredients in evaluating a local transition project. Similarly, national situations may be important in the explanation of success or failure of the national program.

The search for cause has been a debatable issue throughout the history of science, (Mill, 1906; Brodbeck, 1973; Ennis, 1973; Mackie, 1974; and Smith, 1981). Novelists, historians and philosophers have puzzled the nature of cause. In War and Peace, Tolstoy berated historians for predilictions to identify a small number of grand causes--such as Napolean's personality, the length of supply lines, severity of Russian winters, or even a particular supply line in a particular winter. To Tolstoy a corporal burning his breakfast beans on the morning of battle could be as essential a component of victory as his General's indecisiveness. Tolstoy stressed that all that happened was part of the event--all antecedents, actions, and outcomes--and none was more determining than the other.

Gowin repeats this emphasis on "the event" in an earlier chapter. Von Wright (1971) would classify comprehension of the event as "understanding," in contrast to comprehension of causalities, "the explanation." To him the distinction between explanation and understanding underlies key arguments in the philosophy of science and the choice of research methods. The more common target for educational researchers is the causal explanation. Few subscribe to single causes, but for most researchers the gradient from incidental condition to precipitating condition is rather steep. The researcher is expected to concentrate on those variables which have considerable "determining potential." Contexts are studied mainly to learn the limits of causal relationships.

We have two "causalities" to keep track of here: the degree to which something repeatedly creates an effect and the degree to which on a given occasion a contextual circumstance is a principal influence on the outcomes. In the following paragraphs I am emphasizing the latter.

Program evaluators trained in the methods of educational research are likely to search harder for oft-repeating causes than for situational causes. They would like to identify the generalizable cause. Of course some particular conditions persist at an individual site, and thus "generalize over time" locally, even though they may be rare conditions elsewhere. A militancy of the teacher union and the availability of a garden plot are examples. The evaluator recognizes many contexts, some enduring and some momentary, some common and some unique, and decides how much to observe each.

Fixating on common, enduring, (i.e., generalizable) conditions can be a mistake (Cronbach, 1981). It can be an unfortunate slip in role--from program evaluator to educational researcher. As Scriven pointed out (1967), understanding the quality of the particular program often should not be sacrificed to gain general insight into program design. What will serve evaluation sponsors or program stakeholder interests best? Immediate audiences often will be better served by focus on the specific and indifference to broad generalizability. Distant audiences often need help with generalizability to other individual programs. Program evaluators need to understand the interests of their audiences.

A program's situational variables are often identified and tested with analysis of variance. Contexts sometime are singled out for comparison. In evaluating distributive services programs employing handicapped youth, for example, an evaluator may compare school settings with industrial settings as program sites. In Lee Cronbach's APA Distinguished Award address "Beyond the Two Disciplines of Scientific Psychology," he made a grand acknowledgment of context (1973). He reminded us that contextual features peculiar to the site (if they show up at all) will show up in interaction terms.

All the many features and combinations cannot be included in any one analysis. Suppose that in an evaluation of a transition program we have a program director whose immediate prior work was as a Mormon missionary, that the director before him had been fired for general laxity, and that the city's three assembly-line factories had all closed. Is it reasonable to suppose these are extenuating circumstances? Yes. Would they show up in as a significant interaction term in analysis of variance? Probably not. It is only as consistencies in such conditions are found over many programs that interaction variance becomes significant. But sometimes they do.

Even after years of analyses of variance we in educational research have not generated a good understanding of the frequency with which higher order interactions are explanatory. Our intent has been to add additional main effects to designs to provide covariance control and to determine limits of generalization, but not to identify combinations of variables having explanatory value. This particular need for understanding context calls for the first of the three lines of research identified in this paper.

In program evaluation we have opportunity to examine carefully many more contextual aspects than we do in traditional instructional research studies. Our experience as evaluators and our feedback from previous and potential audiences tell us a great deal about which contexts to pay attention to. But our designs should be based on more systematic and rigorous study than has been the case. Direct investigation might be useful but the needed observations are most likely to come as an added ramification to existing projects. Evaluators should give extra attention to such questions as: What aspects of staffing, spatial arrangement, program history and community involvement can be shown to have influence the quality of this program?

Practitioner Generalization.

Now we shift attention to the question of how contextual data aid a reader in deciding whether or not to generalize the findings from the evaluated cases to other cases. I think this is the most important of the three aspects of contextuality raised in this paper.

Each program studied is a case. Sometimes we are primarily interested in this case itself, sometimes in this case as representative of an aggregate of cases. Each of the OSERS funded projects is a case. Evaluation is aimed at each case and at the aggregate of all the cases. Some evaluation designs will be oriented to facilitate generalization to all the cases. But other designs will be oriented to the understanding of this particular case, with expectation sometimes that audiences will make naturalistic generalizations (i.e., extrapolations, applications, expectations based at least partly on personal experience; see Stake and Trumbull, 1982) about this case to other individual cases. Von Wright (1971) made that fundamental distinction between research for generalization and research for particularization. Whether readers seek broad generalization or an understanding of the nature of the particular case, the grounds for forming those generalizations should be understood by the evaluator.

Some authors distinguish between generalizing to populations and extrapolating from one case to another. It is an important distinction, but both are forms of generalization. Application of findings to an additional single case is treated here as "generalization."

Looking for Similarity. Whether the reasoning is formal or intuitive, the basis for generalization is similarity. People generalize more confidently across those things seen similar; they resist generalizing across things seen as different (Strawson, 1952). These similarities and differences pertain to integral features and to the contextual as well. If, for example, program organizations are similar, we are more likely to generalize. And if program contexts are similar, we are more likely to generalize. Urban project directors are more likely to attend to evaluation findings from urban projects.

Of course it is seldom that simple. For real educational programs there are countless features and combinations of features which can be examined for similarity or dissimilarity: the goals, the spirit, the language spoken, the sophistication of the students, the director's qualifications, the facilities for instruction, the commitment to mainstreaming, the availability of a photo-copy machine, the personality of the custodian, and on and on, each of these features itself capable of endless subdivision. *Any* two programs will differ widely on some things, little on others. Yet we somehow manage to feel comfortable that some programs are similar, some are not. In some intuitive way we weight the features, treating a few as determining a similarity, with other features more or less irrelevant (Polanyi, 1958; Glass, 1979; Beyth-Marom and Dekel, 1985).

On what grounds do we select features to be awarded most weight? Kahneman and Tversky (1973) spoke of "representativeness." I think weights are assigned partly in terms of perceived covariation with effect. We increasingly become sensitized to instrumentalities of each feature. A feature is important in terms of its effects; e.g., an employer's provision of instruction for the physically handicapped affects product costs. Whether we treat two employment conditions as similar will depend on the similarity of those features (such as instruction) which affect product cost or other effects of interest to us. Effects won't always be considered. Some weighted features merely covary with other important descriptors. E.g., programs with federal funding have a special place in district organizations whether or not they add to or subtract from productivity. But perceived instrumentality is important.

To find the key features we may ask the question, "Do other things change if this feature is changed?" Take for example a program in a community where people have high "achievement need." We may believe that if aspiration were to diminish it would have an important effect on the program. This then is a contextual feature to which we would assign substantial weight when considering programs as similar.

As evaluators we should be looking for ways programs are seen to be similar. And looking further for bases for considering how findings from the study of one program can be considered relevant to others. Programs studied are bases for understanding *similar* programs. From experience we know something about *which* programs behave the same. From research we know something of consistencies of covariation of common variables. The principle is the same: there is at least some validity to generalizing across similar programs.

The problem of generalization from randomized experiments is addressed by effective assignment of treatments to population samples. When a replicable effect is found we are confident that this effect is of importance in the population of cases. But we do not know that the effect is important for particular individual cases even when they clearly are members of the population. Many educators are mainly concerned about one particular program. Even if experiments are well done, practitioners have to judge whether or not the findings relate to their particular cases. In all kinds of research the findings can be considered generalizable to practice only when practitioners see a basis for generalizing.

Helping Practitioners. The practitioner decides what is relevant. This is not to say that practitioners have better grounds for generalizing than researchers. Actually, intuitive generalizing appears to be an essential aspect of all research, certainly of applied research including program evaluation. The task for us here is to identify grounds for generalization that evaluators can take note of and thereby make their studies more useful (Weylage, 1981).

Consider the evaluation of a hypothetical transition project. A school district has been funded to provide exemplary counseling and job placement for handicapped youth no longer attending school. Early efforts were oriented toward sheltered workshops but more recently efforts have been toward community placement. The change in philosophy apparently has brought about shorter terms of employment and longer terms of unemployment for the students served by the program. Morale of students and staff has dropped. With help from a technical assistance specialist the project has generated extensive records on each student's work history.

The following contextual features are among the facts that the evaluation report might or might not include:

1. In years past there was no transition support for such youngsters in this community.
2. The director has been looking for a new position for a year.
3. Parents of handicapped children have opposed the change in philosophy and are withdrawing voluntary assistance.
4. A popular fundamentalist minister is running for the School Board opposing any contracts for federal programs.

If we knew these situational details, many of us would be more cautious about generalizing that community employment placements is likely to result in shorter placements in our own communities. Would these be typical of the most useful contextual data to obtain? How far afield is it use evaluation funds in the search for bases for reader generalization? We need research that indicates what within our existing evaluation studies is useful in practice.

It is easy to slip into wondering what contextual information helps us explain program workings and impact. That was the main issue of the first section of this paper, an important issue that has traditionally been worked on by educational researchers, yet one that needs continuing study. The question posed here is different: What contextual information helps a practitioner decide whether or not one project's evaluation findings are applicable to other projects?

In the example above the main finding is that adoption of a community placement policy has negatively influenced placement. Should directors of other projects consider the finding pertinent to theirs? Will contextual information help them decide? What contextual information should the evaluator gather for this purpose? Research could help us decide.

It is important for us to remind ourselves that even directors of programs in greatly different contexts may consider the reported covariation between placement policy and the length of the placement to be relevant. For some purposes they might heavily weight size of city or level of unemployment, and see the evaluated program as unlike their own. We all know that some relationships hold over diverse settings. And sometimes a relationship becomes more clear in a highly unusual case (See Simons, 1980).

Were the world a simple place and entirely researched, we could know which relationships hold where and when. As it is, we have no choice but to rely on intuitive sense (Geertz, 1983; Sadler, 1984). But we do not need to be ignorant of the workings of that sense. It is reasonable to suppose that epistemological research on bases for generalization might clarify all uses of evaluation studies. The immediate aim is modest: to understand better how transition program personnel find value in context-rich program evaluation reports.

Evaluator Purview of Context and Emphasis on Program Impact.

Many evaluators have an overly optimistic predisposition toward program design and management. They see programs more available to personal control than they actually are. They attribute failure to exercise control or to accomplish change to poor organization, lack of commitment, inadequate talent or bad luck. The system itself--whatever its sweep--appears to them to be susceptible to management. Contextual constraints are to be subdued or circumvented by clever maneuvers.

On the other hand, many evaluators see program directors and staff members as locked in. Nothing seems to yield to their efforts. Some believe that action at a higher level of governance would unlock the lock, but others are dubious there too. When questioned, few evaluators doubt the possibility of grand change. They see massive changes in history, changes even from year to year, and they see that some of these changes were precipitated by choice, even by rational decision. So only a very few accept the deterministic idea that all these changes are mere evolution of the social system, little influenced by the action of leaders or the collective will of constituents. Such a belief--not uncommon among writers such as Tolstoy--would be admission of the futility of their craft. Most believe that change is possible and manageable.

The activist view, a fundamental sense of "destiny control," is related to the accessibility of mechanisms. In philosophy structuralism attends to matters of life that are deep and inaccessible. Life and life's institutions are usually seen as changing only in superficial ways. As his expression of the indeterminacy principle physicist David Bohm said,

> "No matter how far one goes in the expression of the laws of nature, the result will always depend in an unavoidable way on essentially independent contingencies which exist outside the context under investigation."

The key idea for us now is whether the educational program (figure) is seen to be the pawn or the master of its context (ground). Is the transition program indeed subject to staff control?

According to philosophers such as Saussure (see Georges Mounin, 1986) and enthnographers such as Levi-Strauss, the important things to study are things little subject to personal control. From a structuralist viewpoint such as theirs, an innovative program merely provides a new name and new trappings for what continues to go on. Perceptions of its parts will be altered, some details possibly turned completely around, but with little of essence changing. In this view, any appearance of success of the innovation is achieved by rationalization of purpose and redefinition of evaluative criteria. An appearance of success takes advantage of peripheral manifestation--which, unlike fundamental change, can indeed be produced.

A completely determinist view holds that neither the program itself nor any of the surrounding social systems has the power to make substantive change. But some viewers--not determinists in full sway, but people having a deterministic persuasion--are to be found who hold that few programs have little real potential for change, and that the surrounding social systems do influence the particular program in important ways. Some evaluators with a deterministic persuasion hold human action to be subordinate to a dominating context.

We are unlikely to find any evaluator a complete determinist, nor complete activist. Each evaluator recognizes limit to human power. Each sees at least a few opportunities for control. There is a variety of positions taken along the perspectivial continuum. Every viewer's view is some combination of the possible and the impossible, the doable and the not doable. And the mix will change from setting to setting, and from time to time. Amid this variety and flux it will be seen that some individual evaluators are overly optimistic about change, and some overly pessimistic. And further that evaluation specialists as a group will lean at times too far one way or the other.

Designs. These contrasting purviews have implication for the designing of evaluation studies. If evaluators are too optimistic about the "control potential" held by program personnel and others, they will overdignify the concept of impact and set unrealistic standards for success. In much of Western thinking, impact is the grand concept of program evaluation, not harmony or quality of process or provisioning. A good program is often said to be one that succeeds in producing a desired change. And of course within that global notion impact can be modestly or immodestly defined. The evaluator too optimistic about what can be accomplished performs a disservice to the program by setting an unattainable target.

On the other hand, if evaluators are too pessimistic about the control people have over program destiny, they will devote too little of the design to the discovery of effects. They will tend to draw attention to mere description of process, to the elegance of purpose, to the quality of arrangement, to the covariation of endeavors, and to the intrigue of the story. Accomplishment may be treated as ephemeral and without substance. The evaluator too pessimistic about what can be accomplished performs a disservice to the program by not looking for impact.

For any given program at any given time, the evaluator might be either too optimistic or too pessimistic about impact. Consistently setting unattainable targets or denying the possibility of desired change should be seen as a form of evaluator incompetence. I think we can say that it is a form of unethical practice to persist with such designs in the face of sustained professional and popular complaint. An ethical question probably should be raised also when program evaluation designs uninterruptedly show success, as did years of evaluations of National Science Foundation reform curricula in the 1960s, or designs that uninterruptedly show no success, as for years did the evaluations of Headstart. We understand the influence of evaluator worldview too little. For the meta-evaluation of educational programs we need to set some boundaries on evaluator persuasion toward activism or determinism. We need to be able to spot the evaluator who interprets the aims of program sponsor and staff in such a way that (in the evaluation study) more is required for success than they or any other stakeholders expect to be delivered. Scriven (1973), Eisner (1967), Cronbach and others (1981) have recognized that statements of program goal should not be taken literally as standards of success. Plans reflecting far too robust a view of what people can do should be recognizable by those who evaluate the plans.

But just as clearly we need to spot the evaluator who designs studies that permit the program to escape accountability. No depth of understanding of the intentions, arrangements and complexity of problems is a substitute for an examination of the effects attributable to the program. The evaluator with views pervaded by the "intractability of the system" should be recognized by those who evaluate evaluation plans.

Between these extremes there should be a zone where activism and determinism are sufficiently balanced. (It could be that the range between these two borders is miniscule, or even that some evaluators falling between these bounds are simultaneously too optimistic and too pessimistic, but those conditions are too complicated for this paper). Some would suppose that most evaluators work within this zone, but it is more likely that in the proposals evaluators write, activism predominates. More study is needed to acquaint us with how well professional practices today reflect realistic views of the possibilities of programmatic change.

The continuum here might degenerate into mere tendencies of strictness and leniency to be found in the personality of the evaluator. Personality is important assuredly, but there is an epistemological issue too. It is manifest in the worldview of the evaluator, particularly the degree to which he or she believes critical changes in education lie at the discretion of those responsible for teaching and learning. An alternative way of expressing it is the degree to which he or she believes events are determined by situational factors and systemic structures beyond the reach of the personnel involved. Or in simpler words, the importance of context.

Research Possibilities.

The argument has been threefold, that context is important in evaluative studies but little understood as to its influence (1) on evaluator explanation-making, (2) on reader generalization-making, and (3) on evaluator design-making. One of the first steps in a future research plan of course is to search for existing research along these lines. Few studies are known by this author, but his search has just begun.

Recognizing similarity as a basis for generalization is largely an unconscious process, making it unpromising to just ask report readers what data they use. We need to observe the generalizations from actual evaluation reports to discern how features seemed to be weighted. One also could contrive evaluation reports with various contextual features, perhaps in a balanced design, to learn what features are considered contributors to meaning and application. But these contrived situations are usually too simplistic. The better studies will probably be those providing observations of what people do when exposed to context-rich and context-poor findings.

To understand and improve evaluation practice it is important that the perceived roles of context be examined. Evaluation specialists should work with epistemologists, drawing from psychological studies of perception, hermeneutics, and the nature of evidence--logical, legal, and popular. Proposals of evaluation studies, final reports, and formal and informal statement of evaluators should be examined. We of course cannot get directly at the thinking processes of evaluators and evaluation report users. Our investigation of generalization power and the utility of contextual data can only be indirect. But thorough study of our own work should give us a greater sense of which observations and findings become useful and which do not.

References

Beyth-Marom, R. and Shlomith D. (1985). An elementary approach to thinking under uncertainty. Hillsdale, NJ: Lawrence Erlbaum.

Bohm, D. (1957). Causality and chance in modern physics. New York: Harper.

Brandt, R.M. (1972). Studying behavior in natural settings. NY: Holt.

Brodbeck, M. (1973). "Causation." In H.S. Broudy, R.H. Ennis, and L.I. Krimerman (Eds.), Philosophy of educational research. NY: Wiley.

Cronbach, L.J. (1975, February). "Beyond the two disciplines of scientific psychology." American Psychologist, 30, 116-127.

Cronbach, L.J. and Associates. (1981). Toward reform of program evaluation. San Francisco: Jossey-Bass.

Cronbach, L.J. (1982). "Prudent aspirations for social inquiry." In W. Kruskal (Ed.), The state of the social sciences: Fifty years at Chicago. (pp. 61-81). Chicago: University of Chicago Press.

Eisner, E.W. (1977). "Instructional and expressive objectives: their formulation and use in curriculum. AERA Monograph Series on Curriculum Evaluation #3. Chicago: Rand McNally.

Ennis, R.H. (1973, June). "On causality." Educational Researcher, 2, 6.

Erickson, F. (1986). Qualitative methods in research on teaching. In M.C. Wittrock, (Ed.), Handbook of research on teaching, (3rd ed.). NY: Macmillan.

Geertz, C. Local knowledge. NY: Basic Books, 1983.

Glass, G.V. (1979). "Policy for the unpredictable." Educational Researcher, 8(9), 12-14.

Guba, E.G. & Lincoln, Y. (1981). Effective evaluation. San Francisco: Jossey-Bass.

Kahneman, D. & Tversky, A. (1973). On the psychology of prediction. Psychological Review, 80, 237-251.

Levi-Strauss, C. (1955). Tristes tropiques. Paris: Plon.

Mackie, J.L. (1974). The cement of the universe: a study of causation. Oxford: Clarenden Press.

Mill, J.S. (1906). A system of logic, I. London: Longman's, Green.

Mounin, G. (1968). Saussure, ou le structuralisme sans le savoir. Paris: Hatiers.

Parlett, M. & Hamilton, D. (1976). "Evaluation as illumination." In G.V. Glass (Ed.), Evaluation studies review annual (Vol. 1, pp. 140-157). Beverly Hills: SAGE.

Patton, M.Q. (1981). Creative evaluation. Beverly Hills: SAGE.

Pirsig, R.M. (1974). Zen and the art of motorcycle maintenance. New York: Wm. Morrow & Company.

Polanyi, M. (1958). Personal knowledge. Chicago: University of Chicago Press.

Sadler, D.R. (1984, April). Evaluation and the logic of recommendations. Evaluation Review, 8(2), 261-268.

Schatzman, L. & Strauss A. (1973). Field research: strategies for a natural sociology. Englewood Cliffs: Prentice Hall.

Schutz, A. (1963). Common-sense and scientific interpretation of human action. In M. Natanson (Ed.), Philosophy of the social sciences. New York: Random House.

Scriven, M. (1973a). "Goal-free evaluation." In Ernest R. House (Ed.), School evaluation: the politics and process. Berkeley: McCutchan Publishing Corporation.

Simons, H. (Ed.). (1980). Toward a science of the singular. Norwich, England: CARE, UEA.

Smith, N. (1981, March). Noncausal inquiry in education. Educational Researcher.

Stake, R.E., Easley, J. and others. (1978). Case Studies in Science Education. Champaign-Urbana: University of Illinois, Center for Instructional Research and Evaluation.

Stake, R.E. (1981). "Case study methodology: an epistemological advocacy." In W.W. Welch (Ed.), Case Study Methodology in Educational Evaluation. Proceedings of the 1981 Minnesota Evaluation Conference, U of Minn.

Stake, R.E. & Trumbull, D.J. (1982). Naturalistic generalizations. Review Journal of Philosophy and Social Science, VII, (1 & 2).

Strawson, P.F. (1952). Introduction to logical theory. London: Methuen.

Von Wright, G.H. (1971). Explanation and understanding. London: Routledge and Kegan Paul.

Weylage, G. (1981). The purpose of generalization in field-study research. In T.S. Popkewitz and B.R. Tabachnick, (Eds.), The study of schooling: field based methodologies in educational research and evaluation. New York: Praeger.

Wolcott, H. (1976). Criteria for an ethnographic approach to research in schools. In J.T. Roberts and S.K. Akinsanya, (Eds.), Schooling in the cultural context. NY: David McKay.

BIBLIOGRAPHY FOR EVALUATIVE RESEARCH ON TRANSITION PROJECTS

Agarwala-Rogers, R. (1977). Why is evaluation research not utilized? In M. Guttentag (Ed.), *Evaluation Studies Review Annual* (*2*, 327-333). Beverly Hills, CA: Sage Publications.

Alkin, M. C. (1975). Evaluation: Who needs it? Who cares? *Studies in Education Evaluation*, *1*, 201-212.

Alkin, M. C. (April 1980). *Can we improve technical assistance: Implications of utilization research*. Paper presented at the American Educational Research Association annual meeting, Boston, MA.

Alkin, M. C., & Daillak, R. H. (1979). A study of evaluation utilization. *Educational Evaluation and Policy Analysis*, *1*(4), 41-49.

Alkin, M. C., Daillak, R., & White, P. (1979). *Using evaluations: Does evaluation make a difference*? Beverly Hills, CA: Sage Publications.

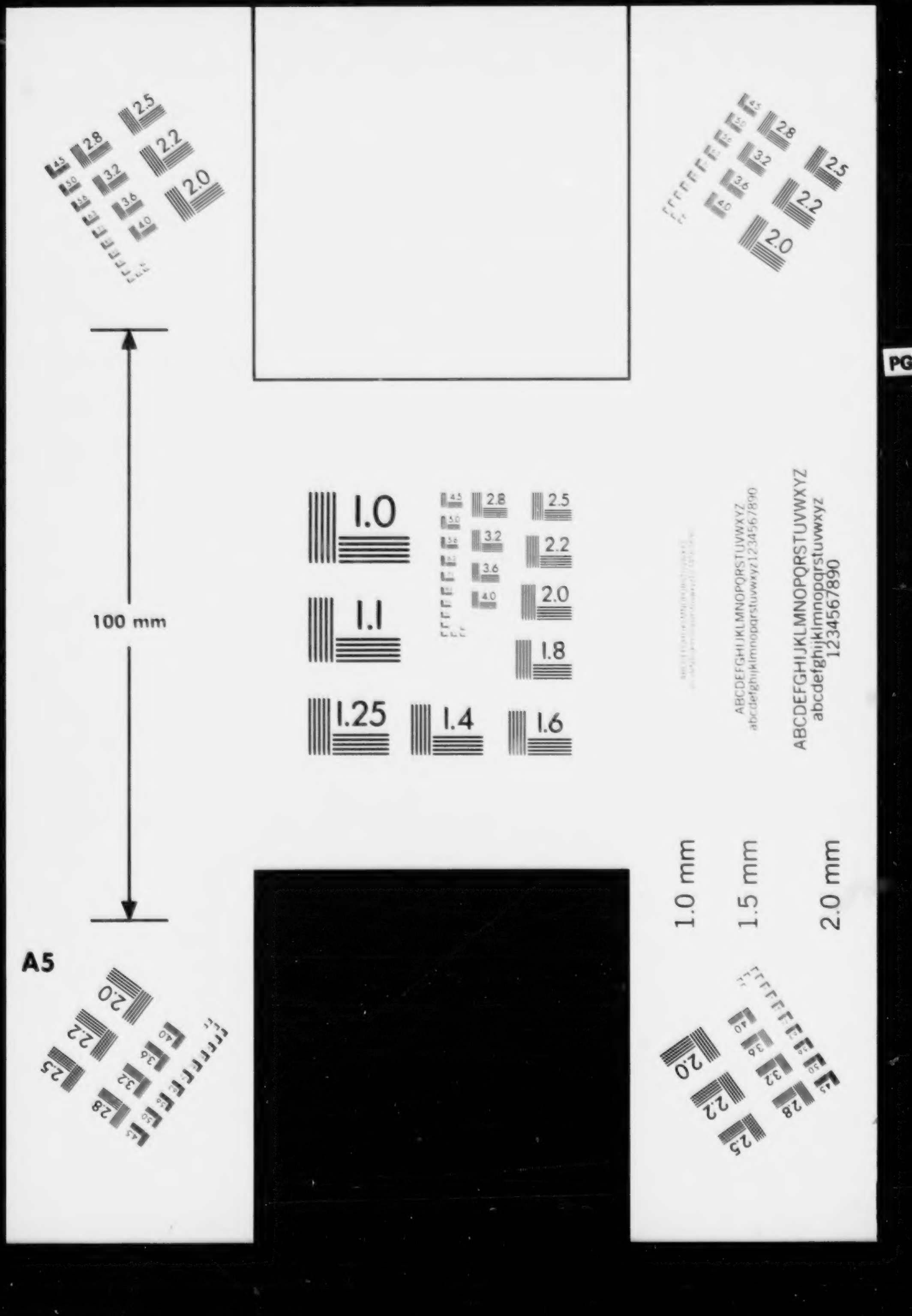
Alkin, M. C., Kosecoff, J., Fitz-Gibbon, C., & Seligman, R. (1974). Evaluation and decision making: The Title VII experience. *CSE Monograph Series in Evaluation*, *4*.

Anderson, S. B. (1975). *Encyclopedia of educational evaluation*. San Francisco: Jossey-Bass.

Anderson, S. B., & Ball, S. (1978). *The profession and practice of program evaluation*. San Francisco: Jossey-Bass.

Apple, M., Subkoviak, J., & Lufler, H. S. Jr. (Eds.). (1974). *Education evaluation: Analysis and responsibility*. Berkeley, CA: McCutchan Pub. Corp.

4


PG
100 mm
A5
1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
5.6
ABCDEFGHIJKLMNOPQRSTUVWXYZ
abcdefghijklmnopqrstuvwxyz1234567890
ABCDEFGHIJKLMNOPQRSTUVWXYZ
abcdefghijklmnopqrstuvwxyz
1234567890
1.0 mm
1.5 mm
2.0 mm

Archibald, K. A. (1970, April). Alternative orientations to social science utilization, *Social Science Information*, *9*(2), 7-34.

Arieh, L., & Nevo, D. (Eds.). (1981). *Evaluation roles in education*. London: Gordon and Breach.

Arrow, K. (1974). *The limit of organization*. New York: Norton.

Asher, J. W. (1976). *Education research and evaluation methods*. Boston, MA: Little, Brown.

Aslanian, C. B. (1981). *Improving educational evaluation methods*. Beverly Hills, CA: Sage Publications.

Attkisson, C. C. (Ed.). (1978). *Evaluation of human service programs*. New York: Academic Press.

Ball, S., & Anderson, S. B. (1977). Dissemination, communication, and utilization. *Education & Urban Society*, IX, 451-470.

Baratz, S. S., & Marvin, K. E. (1978). Resolving privacy, access, and other problems in the audit and reanalysis of social research for policy. *Evaluation and Change*, 31-35.

Bardach, E. (1976). Policy termination as a political process. *Policy Sciences*, *7*, 123-131.

Barnow, B. S., & Cain, G. G. (1977). A reanalysis of the effect of head start on cognitive development: Methodology and empirical findings. *Journal of Human Resources*, *12*, 177-197.

Barton, P. E., & Fraser, B. S. (1978). *Between two worlds: Youth transition from school to work* (Vols. 1-3). Washington, DC: Center for Education and Work, National Manpower Institute.

Behn, R. D. (1978). How to terminate a public policy: A dozen hints for the would be terminator. *Policy Analysis*, *4*, 393-413.

Beigel, A. (1974). Evaluation on a shoestring. In Wm. A. Hargreaves, et al. (Eds.), Resource materials for community mental health program evaluation. San Francisco: NIMH.

Benz, M., & Halpern, A. (in press). A statewide examination of secondary special education for students with mild disabilities: Implications for transition services. Exceptional Children.

Benz, M., & Halpern, A. (1986). Vocational preparation for high school students with mild disabilities: A statewide study of administrator, teacher, and parent perceptions. Career Development for Exceptional Individuals, 9, 3-15.

Berke, I. P. (1979, October). Utilization of educational program evaluation for federal policy making: The case of bilingual education. Paper presented at the Evaluation Research Society annual meeting, Minneapolis, MN.

Bennett, R. E. (1981). Evaluating individualized education programs. Diagnostique, 7(2), 91-100.

Bickman, L., & Weatherford, D. L. (Eds.). (1986). Evaluating early intervention programs for severely handicapped children and their families. Austin, TX: Pro-Ed.

Blanton, J., & Alley, S. (1978). How evaluation findings can be integrated into program decision making. Community Mental Health Journal, 14(3), 239-247.

Boruch, R. F., & Gomez, H. (1979). Measuring impact: Power in social program evaluation. In L. E. Datta & R. Perloff (Eds.), Improving evaluations. Beverly Hills, CA: Sage Publications.

Braskamp, L. A., & Brown, R. D. (Eds.). (1980). Utilization of evaluative information. New Directions for Program Evaluation, (5).

Braskamp, L. A., Brown, R. D., & Newman, D. L. (1978). Credibility of a local educational program evaluation report: Another source and client characteristics. American Educational Research Journal, 15(3), 441-450.

Brickell, H. M. (1974). Data for decisions: An analysis of evaluation data needed by decision-makers in educational programs. New York: Policy Studies in Education.

Brictson, P. (1980, April). What evaluation data are useful to legislatures and state boards of education? Paper presented at the American Educational Research Association annual meeting, Boston, MA.

Brown, R. D., & Braskamp, L. A. (Eds.). (1980). Utilization of evaluative information. San Francisco: Jossey-Bass.

Brown, R. D., Braskamp, L. A., & Newman, D. L. (1978). Evaluator credibility as a function of report style: Do jargon and data make a difference? Evaluation Quarterly, 2(2).

Bronfenbrenner, U. (1979). The ecology of human development: Experiments by nature and design. Cambridge, MA: Harvard University Press.

Campbell, D. T. (1974). Qualitative knowing in action research. Occasional paper, Stanford Evaluation Consortium, Stanford, CA.

Campbell, D. T. (1969b). Reform as experiments. American Psychologist, 24, 409-429.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin, 56, 81-105.

Campbell, D. T., & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed.), Handbook of research on teaching. Chicago: Rand McNally.

Caplan, N. (1977). A minimal set of conditions necessary for the utilization of social science knowledge in policy formation at the national level. In C. H. Weiss (Ed.), Using social research in public policy making, (pp. 183-197). Lexington, MA: Lexington Heath.

Caplan, N. (1977). Social research and national policy: What gets used by whom, for what purposes, and with what effects. In M. Guttentag (Ed.), Evaluation studies review annual, (Vol. 2). Beverly Hills: Sage Publications.

Carlson, Paul. (1973). The relationship of three organizational variables of the utilization of evaluation. Unpublished doctoral dissertation, Faculty of Educational Development, Ohio State University.

Carter, R. K. (1973). Clients' resistance to negative findings. In E. R. House (Ed.), School evaluation: The politics and process. Berkeley: McCutchan.

Chelimsky, E. (Ed.). (1977, March). Proceedings of a symposium on the use of evaluation by federal agencies, (Vol. 2). Washington, D.C.: Mitre Corporation (M77-39).

Cherney, P. R. (Ed.). (1971). Making evaluation research useful. Columbia, MD: American City Corporation.

Ciarlo, J. A. (Ed.). (1981). Utilizing evaluation: Concepts and measurement techniques. Beverly Hills, CA: Sage Publications.

Cicchinelli, L. F. (1977). Assessment of state and local government evaluation practices: An evaluation unit profile. Austin, TX:

Program Evaluation and Review Division, Texas Department of Public Welfare.

Close, D. W., & Halpern, A. S. (1984). Measuring the learning potential of severely mentally retarded adults. In A. S. Halpern & M. J. Fuhrer (Eds.), *Functional assessment in rehabilitation* (pp. 159-170). Baltimore: Paul H. Brookes Publishing Co.

Close, D. W., Sowers, J., Halpern, A. S., & Bourbeau, P. (1985). Programming for the transition to independent living for mildly retarded persons. In K. C. Lakin & R. H. Bruininks (Eds.), *Strategies for achieving community integration of developmentally disabled citizens* (pp. 161-176). Baltimore: Paul H. Brookes Publishing Co.

Cohen, B. F., & Anthony, W. A. (1984). Functional assessment in psychiatric rehabilitation. In A. S. Halpern & M. J. Fuhrer (Eds.), *Functional assessment in rehabilitation* (pp. 78-100). Baltimore: Paul H. Brookes Publishing Co.

Cohen, D. K., & Farrar, E. (1977). Power to the parents? - The story of education vouchers. *The Public Interest*, (48), 72-97.

Cohen, D. K., & Garet, M. S. (1975). Reforming educational policy with applied social research. *Harvard Educational Review*, *45*(1), 17-43.

Cohen, L. H. (1977). Factors affecting the utilization of mental health evaluation research findings. *Professional Psychology*, *8*(4), 526-634.

Coleman, J. S. (1972). *Policy research in the social sciences*. Morristown, N.J.: General Learning Press.

Coleman, J. S., et al. (1966). Equality of educational opportunity. Washington, DC: U.S. Government Printing Office.

Coles, C. D., & Anderson, S. B. (Eds.). (1978). Exploring purposes and dimensions. San Francisco: Jossey-Bass.

Comptroller General. (1977). Problems and needed improvements in evaluating office of education programs. Washington, DC: U.S. Government Printing Office.

Conner, R. F. (1979). The evaluation of research utilization. In M. W. Klein and K. S. Teihmann (Eds.), The handbook of criminal justice evaluation, Beverly Hills, CA: Sage Publications.

Cook, T. D., & Campbell, D. T. (1979). Quasi-experimentation: Design and analysis issues for field settings. Chicago: Rand McNally.

Cook, T. D., & Pollard, W. E. (1977). Guidelines: How to recognize and avoid some common problems of mis-utilization of evaluation research findings. Evaluation, 4, 161-164.

Cooley, W. W., & Lonnes, P. R. (1976). How can evaluations influence decisions and policies? In W. W. Cooley and P. R. Lohnes (Eds.), Evaluation research in education: Theory, principles, and practice (pp. 339-350). New York: Wiley and Sons.

Coulson, J. E. (1980, April). Making national evaluations more acceptable to schools: An evaluation contractor's perspective. Paper presented at the American Educational Research Association annual meeting, Boston, MA.

Cox, G. B. (1977). Managerial style implications for the utilization of program evaluation. Evaluation Quarterly, 1(3), 499-500.

Crandall, D. P., & Harris, R. C. (1978, September). Views on the utilization of information by practitioners. Andover, MA: The Network.

Cronbach, L. J. (1975). Beyond the two disciplines of scientific psychology. American Psychologist, 30, 116-117.

Cronbach, L. J. (1980). Toward reform of program evaluation. San Francisco: Jossey-Bass.

Cronbach, L. J. (1982). Designing evaluations of educational and social programs. San Francisco: Jossey-Bass.

Cronbach, L. J., & Snow, R. E. (1977). Aptitudes and instructional methods. New York: Irvington.

Crystal, R. M., & Lee, C. C. (Eds.). (1979). Program evaluation terms: A glossary (Michigan Studies in Rehabilitation Utilization Series: 3). Ann Arbor: Michigan University, Rehabilitation Research Institute.

Datta, L. (1976). The impact of the Westinghouse/Ohio evaluation of the development of project Head Start: An examination of the immediate and longer-term effects and how they came about. In C. C. Abt (Ed.), The evaluation of social programs (pp. 129-181), Beverly Hills, CA: Sage Publications.

Datta, L. (1977). Does it work when it has been tried? And half full or half empty? In M. Guttentag (Ed.), Evaluation studies review annual, (Vol. 2), Beverly Hills, CA: Sage Publications, 1977, 301-319.

David, J. L. (1978). Local uses of Title I evaluations. (Report EDRC 21, SRI Project URU-6854). Menlo Park, CA: SRI International.

Davis, H. R., & Salasin, S. E. (1975). The utilization of evaluation. In E. L. Struening and M. Guttentag (Eds.), *Handbook of evaluation research*, (Vol. 1, pp. 621-666). Beverly Hills, CA: Sage Publications.

Deal, T. R., & Huguenin, K. M. (1975, August). *Water and the duck's back: The use of formative evaluation in schools*. Stanford, CA: Stanford University Center for Research and Development in Teaching. (ERIC Document Reproduction Service No. Ed 134 608)

Dexter, L. A. (1966). Impressions about utility and wastefulness in applied social science studies. *American Behavioral Scientist*, *9*(6), 9-10.

Dunst, C. J. (1979). Program evaluation and the education for all handicapped children act. *Exceptional children*, *46*(1), 24-31.

Early, R. (1980, April). *What evaluation data are useful to teachers?* Paper presented at the American Educational Research Association annual meeting, Boston, MA.

Elmer, V. (1979, September). *Strengths and weaknesses of an in-house evaluation unit: (Or, how to succeed by stepping on everyone's toes)*. Paper presented at the Evaluation Network annual meeting, Cincinnati, OH.

Falvey, M. (1985). *Community-based curriculum: Instructional strategies for students with severe handicaps*. Baltimore: Paul H. Brookes Publishing Co.

Fishbein, R. L. (1980, April). *Increasing the relevancy of evaluation data to school practice: The Title I Evaluation Technical Assistance Centers*. Paper presented at the American Educational Research Association annual meeting, Boston, MA.

Fisher, D. G. (1984). *The ethological program evaluation method*. Unpublished doctoral dissertation, University of Illinois, Champaign-Urbana.

Frey, W. D. (1984). Functional assessment in the '80s. In A. S. Halpern & M. J. Fuhrer (Eds.), *Functional assessment in rehabilitation* (pp. 11-43). Baltimore: Paul H. Brookes Publishing Co.

Galagan, J. (1985). Psychoeducational testing: Turn out the lights, the party's over. *Exceptional Children*, *52*, 288-299.

Gilbert, J. P., Light, R. J., & Mosteller, F. (1975). Assessing social innovations: An empirical base for policy. In C. A. Bennett & A. A. Lumsdaine (Eds.), *Evaluation and experiment*. New York: Academic Press.

Gill, L. (1979, October). *Reporting strategies for the use of evaluation as a problem solving aid*. Paper presented at the Evaluation Research Society annual meeting, Minneapolis, MN.

Ginzberg, E. (1980). *The school/work nexus*. Bloomington, IN: Phi Delta Kappa Educational Foundation.

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago: Aldine.

Glass, G. V. (1976). Primary, secondary and meta analysis of research. *Educational Researcher*. *5*(11), 3-8.

Glass, G. V. (1979). Policy for the unpredictable (uncertainty research and policy). *Educational Researcher*. *8*(9), 12-14.

Gold, N. (1979, October). *Experience with the stakeholder: Utilization strategy with the cities, in schools, and push for excellence programs*. Paper presented at the Evaluation Research Society annual meeting, Minneapolis, MN.

Goldstein, M. S., Marcus, A. C., & Rausch, N. P. (1978). The non-utilization of evaluation research. Pacific Sociological Review, 21, 21-44.

Granville, A. (1977). Where do school decisions really come from? Unpublished doctoral dissertation, University of California at Los Angeles.

Guba, E. G. (1969). The failure of educational evaluation. Educational Technology, 9(5), 29-38.

Guba, E. G. (1975). Problems in utilizing the results of evaluation. Journal of Research and Development in Education, 8(3), 42-54.

Guba, E. G. (1978). Toward a methodology of naturalistic inquiry in educational evaluation. Los Angeles: Center for the Study of Evaluation, UCLA Graduate School of Education.

Guba, E. G., & Lincoln, Y. (1981). Effective education. San Francisco: Jossey-Bass.

Haenn, J. (1980, April). Reasons why testing and evaluation don't inform. Paper presented at the American Educational Research Association annual meeting, Boston, MA.

Halpern, A. (1974). Work-study programs for the mentally retarded: An overview. In P. Browning (Ed.), Mental retardation: Rehabilitation and counseling. Springfield, IL: Charles C Thomas.

Halpern, A. (1984). Functional assessment and mental retardation. In A. Halpern & M. Fuhrer (Eds.), Functional assessment in rehabilitation (pp. 61-78). Baltimore: Paul H. Brookes Publishing Co.

Halpern A., & Benz, M. (in press). A statewide examination of secondary special education students with mild disabilities: Implications for the high school curriculum. Exceptional Children.

Halpern, A., Close, D., & Nelson, D. (1986). On my own: The impact of semi-independent living programs on adults with mental retardation. Baltimore: Paul H. Brookes Publishing Co.

Halpern, A. S., & Fuhrer, M. J. (1984). Functional assessment in rehabilitation. Baltimore: Paul H. Brookes Publishing Co.

Halpern, A., Lehmann, J., Irvin, L., & Heiry, T. (1982). Contemporary assessment for mentally retarded adolescents and adults. Baltimore: University Park Press.

Halpern, A. Nave, G., Close, D., & Nelson, D. (in press). An empirical analysis of the dimensions of community adjustment for adults with mental retardation in semi-independent living programs. Australia and New Zealand Journal of Developmental Disabilities.

Halpert, H. P. (1969). Communications as a basic tool in promoting utilization of research findings. In H. Shoulberg, A. Sheldon, & F. Kaker (Eds.), Program evaluation in the health fields (pp. 497-505). New York: Behavioral Publications.

Hamilton, D. (1976). Curriculum evaluation. London: Open Books.

Hamilton, D., et al. (Eds.). (1978). Beyond the numbers game. Berkeley, CA: McCutchan.

Haney, W. A. (1977). A technical history of the national follow through evaluation. Cambridge, MA: Huron Institute.

Hastings, J. T. (1966). The why of the outcomes. Journal of Educational Measurement, 3, 27-32.

Hauser, C. (1979). Evaluating mainstream programs: Capitalizing on a victory. Journal of Special Education, 13(2), 107-29.

Hawkins, J. D., Roffman, R. A., & Osborne, P. (1978). Decision makers' judgments: The influence of role, evaluative criteria, and information access. Evaluation Quarterly, 2(3), 435-454.

Hayman, J., Rayder, N., Stenner, A. J., & Madey, D. C. (1979). On aggregation, generalization, and utility in educational evaluation. Educational Evaluation and Policy Analysis, 1(4), 31-39.

Heath, R. W. & Brandon, P. R. (1982). An alternative approach to the evaluation of educational and social programs. Educational Evaluation and Policy Analysis, 10(4), 477-486.

Hobbs, N. (1975). The futures of children: Recommendations of the project on classification of exceptional children. San Francisco: Jossey-Bass.

House, E. R. (Ed.). (1986). New directions in education evaluation. Philadelphia, PA: The Falmer Press.

House, E. R., et al. (1978). No simple answer: Critique of the follow through evaluation. Harvard Educational Review, 48, 128-160.

Ingison, L. J. (1979, October). Some realities and constraints regarding the use and usefulness of policy-relevant evaluations. Paper presented at the Evaluation Research Society annual meeting, Minneapolis, MN.

Isaac, S. (1980, April). Increasing the relevance of national evaluation to schools: A school district's viewpoint. Paper presented at the American educational Research Association annual meeting, Boston, MA.

James, B. N. (1980, April). *Why have evaluations of staff development centers been of limited use so far?* Paper presented at the American Educational Research Association annual meeting, Boston, MA.

Jones, R. L., & others. (1978). Evaluating mainstreaming programs: Models, caveats, considerations and guidelines. *Exceptional Children*, *44*(8), 588-601.

Katzenmeyer, C. G. (1979, September). *Congressional hearings as a model for understanding how evaluations are used for policy and planning*. Paper presented at the Evaluation Network annual meeting, Cincinnati, OH.

Kirst, M. (1978). Short- and long-range consequences of proposition 13. *California School Board Journal*, *37*(6), 27-30.

Knorr, K. D. (1977). Policymakers' use of social science knowledge: Symbolic or instrumental? In C. H. Weiss (Ed.), *Using social research in public policy making* (pp. 165-182). Lexington, MA: Heath.

Koretz, D. M., & Saxe, L. (Eds.). (1982). *Making evaluation research useful to congress*. San Francisco: Jossey-Bass.

Lange, R. R., & Speiss, M. (1974, April). *An empirical study of four providing variables affecting utilization of evaluation*. Paper presented at the American Educational Research Association annual meeting, Chicago.

Levin, H. M. (1975). Cost effectiveness analysis in evaluation research. In M. Guttentag & E. L. Struening (Eds.), *Handbook of evaluation* (Vol. 2). Beverly Hills, CA: Sage Publications.

Levitan, S. K., & Wurgburg, G. (1979). *Evaluating federal social programs: An uncertain art*. Kalamazoo, MI; W. E. Upjohn Institute.

Lindblom, C. E., & Cohen, D. K. (1979). _Usable knowledge: Social science and social problem solving_. New Haven, CT: Yale University Press.

Locatis, C. N., Smith, J. K., & Blake, V. L. (1979, April). _Effects of evaluation information on decisions_. Paper presented at the American Educational Research Association annual meeting, San Francisco.

Longood, R., & Simmel, A. (1972). Organization resistance to innovation suggested by research. In C. H. Weiss (Ed.), _Evaluating action programs: Readings in social action and education_. Boston: Allyn and Bacon.

Lorenzen, G. L., & Braskamp, L. A. (1978). Comparative influences of political cost benefit and statistical evaluation information in administrative decision making. _Evaluation and Program Planning_, _1_(3), 235-238.

Lumsdaine, A. A., & Bennett, C. A. (1975). Assessing alternative conceptions of evaluation. In C. A. Bennett and A. A. Lumsdaine (Eds.), _Evaluation and experiment, some critical issues in assessing social programs_. New York: Academic Press.

Lynd, R. S. (1939). _Knowledge for what? The place of social science in American culture_. Princeton, N.J.: Princeton University Press.

MacDonald, B., & Parlett, M. (1973). Rethinking evaluation: Notes from the Cambridge conference. _Cambridge Journal of Education_, _3_, 74-82.

Mackie, J. L. (1974). _The cement of the universe: A study of causation_. Oxford: Clarendon Press.

Maher, C. A., & Bennett, R. E. (1984). *Planning and evaluating special education services*. Englewood Cliffs: Prentice-Hall, Inc.

Mallory, B. L. (1979). *The ecology of implementation: A qualitative approach to education evaluation*. Ann Arbor: George Peabody College for Teachers.

Mann, J. (1972). The outcome of evaluation research. In C. H. Weiss (Ed.), *Evaluating action research: Readings in social action and education*. Boston: Allyn and Bacon.

March, J. G., & Simon, H. W. (1958). *Organizations*. New York: Wiley.

McLaughlin, M. W. (1975). *Evaluation and reform: The elementary and secondary education act of 1965*. Cambridge, MA: Ballinger.

McPartland, J. M., Dawkins, R. L., & Braddock II, J. H. (1986). *The school's role in the transition from education to work: Current conditions and future prospects* (Rep. No. 362). Baltimore: The John Hopkins University, Center for Social Organization of Schools.

Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology*, *46*, 806-834.

Mercer, J. (1983). *Labeling the mentally retarded*. Berkeley: University of California Press.

Mulliken, R. K., & Buckley, J. J. (1983). *Assessment of multihandicapped and developmentally disabled children*. Rockville, MD: Aspen Systems Corporation.

Murphy, S. T., & Ursprung, A. (1983, February). The politics of vocational education. *Rehabilitation Literature*, *44*(1), 2-12.

Mushkin, E. (1973). Evaluations: Use with caution. Evaluation, 1(2), 31-35.

Nagel, S. & Neef, M. (Eds.). (1975). The political science utilization directory. Urbana, Illinois: Policy Studies Organization.

Neilsen, V. G. (1975). Why evaluation does not improve program effectiveness. Policy Studies Journal, 3(4), 385-390.

Nevo, D., & Stufflebeam, D. L. (1976). The availability and importance of evaluation information within the school. Studies in Educational Evaluation, 2(3), 203-210.

Newman, D. L., Brown, R. D., & Littman, M. (1979). Evaluator report and audience characteristics which influence the impact of evaluation reports: Does who says what to whom make a difference? CEDR Quarterly, 12(2), 14-18.

Odom, S. L., & Fewell, S. R. (1983). Program evaluation in early childhood special education: A meta-evaluation. Educational Evaluation and Policy Analysis. 10(4), 445-460.

Parlett, M., & Hamilton, D. (1978). Evaluation and illumination: A new approach to the study of innovatory programmes. In D. Hamilton, et al. (Eds.), Beyond the numbers game. Berkeley, CA: McCutchan.

Patton, M. Q. (1978). Utilization-focused evaluation. Beverly Hills, CA: Sage Publications.

Patton, M. Q., Grimes, P. S., Guthrie, K. M., Brennan, N. J., French, B. D., & Blyth, D. A. (1978). In search of impact: An analysis for the utilization of federal health evaluation research. In T. D. Cook, et al., (Eds.), Evaluation studies review annual (Vol. 3, pp. 59-80). Beverly Hills, CA: Sage Publications.

Pelz, R. (Ed.). (1982). Evaluation case studies. (WESTAR Series Paper #16). Monmouth, OR: Western States Technical Assistance Resource. Special Education Programs, Washington, DC: Handicapped Children's Early Education Program.

Perkins, D. (1979, October). Social intervention evaluations and public policy. Paper presented at the Evaluation Research Society annual meeting, Minneapolis, MN.

Perloff, R., & Perloff, E. (Eds.). (1980). Values, ethics and standards in evaluation. San Francisco: Jossey-Bass.

Poliuka, L., & Steg, E. (1978). Program evaluation and policy development. Evaluation Quarterly, 2(4), 696-707.

Popper, K. R. (1957). The poverty of historicism. London: Routledge & Kegan Paul.

Posante, R. (1981, April). Utilization of evaluation results in joint policy making. Paper presented at the annual meeting of the American Educational Research Association, Los Angeles, CA.

Prasse, D. P. (1980). School psychology, special education and program evaluation: Ethical and professional concerns. School Psychology Review, 9(3), 267-272.

Rein, M. (1976). Social science and public policy. New York: Penguin Books.

Rein, M., & Schon, D. A. (1977). Problem setting in policy research. In C. H. Weiss (Ed.), Using social research in public policy making. Lexington, MA: Lexington Books.

Rich, R. F. (1977). Uses of social science information by federal bureaucrats: Knowledge for action versus knowledge for understanding. In C. H. Weiss (Ed.), Using social research in public policy making (pp. 199-211). Lexington, MA: Heath.

Ricks, F. A. (1980). Evaluation in Special Education: If you care, you count. B. C. Journal of Special Education, 4(3), 241-244.

Rivlin, A. M. (1971). Systematic thinking for social action. Washington, D.C.: Brookings Institution.

Rivlin, A. M., & Timpane, P. M. (Eds.). (1975b). Planned variation in education. Washington, DC: Brookings Institution.

Rossi, P. H. (Ed.). (1982). Standards for evaluation practice. San Francisco: Jossey-Bass.

Rumbaugh, S. A. (1980, April). Some possible procedures for reducing burden and increasing the relevance of evaluation data: A state perspective. Paper presented at the American Educational Research Association annual meeting, Boston, MA.

Rusch, F. R. (1986). Competitive employment issues and strategies. Baltimore: Paul H. Brookes Publishing Co.

Russo, R. P. (1980). Toward understanding and evaluating special needs programs. Minneapolis: Minnesota University, Department of Vocational and Technical Education.

Salasin, S. (1978). Harings on the cost, management and utilization of human resources program evaluation: Summaries of the testimony. Evaluation, Special Issue, 10-18.

Scriven, M. (1967). The methodology of evaluation. In R. E. Stake, et al. (Eds.), Perspectives on curriculum evaluation. AERA Monograph Series on Curriculum Evaluation, (1). Chicago: Rand McNally.

Simon, H. A. (1976). From substantive to procedural rationality. In. S. J. Latsis (Eds.), Method and appraisal in economics. Cambridge: Cambridge University Press.

Smith, N. L. (1981). Federal efforts to develop new evaluation methods. San Francisco: Jossey-Bass.

Speiss, M. (1973). The effect of four communication variables on the utilization of evaluation conclusions in decision making. Unpublished doctoral dissertation, Faculty of Educational Development, Ohio State University.

St. Pierre, R. G. (Ed.). (1983). Management and organization of program evaluation. San Francisco: Jossey-Bass.

Stalford, C. B. (1979, April). The genesis of a stakeholder strategy for evaluation utilization. Paper presented at the Evaluation Research Society annual meeting, Minneapolis, MN.

Stanford, R. N. (1970). Whatever happened to action research? Journal of Social Issues. 26(4), 3-23.

Stieglitz, M. N., & Cohen, J. S. (1980). Career education for physically disabled students: A bibliography. Washington, DC: National Institute of Handicapped Research.

Stokes, T. & Baer, D. (1977). An implicit technology of generalization. Journal of Applied Behavior Analysis, 10, 349-367.

Straughan, R., & Wrigley, J. (Eds.). (1980). Values and evaluation in education. London: Harper and Row.

Tharp, R. G., & Gallimore, R. (1979). The ecology of program research and evaluation: A model of evaluation succession. In L. B. Sechrest, et al. (Eds.), Evaluation studies review annual, 4 (pp. 39-60).

Trauth, J. (1979, October). The use of evaluation to provide short-term feedback and to suggest early program modifications. Paper presented at the Evaluation Research Society annual meeting, Minneapolis, MN.

Vash, C. (1984). Evaluation from the client's point of view. In A Halpern & M. Fuhrer (Eds.), Functional assessment in rehabilitation. Baltimore: Paul H. Brookes Publishing Co.

Walker, D. F., & Schaffarzick, J. (1974). Comparing curricula. Review of Educational Research, 44, 83-111.

Wallace, G., & Larsen, S. (1972). Educational assessment of learning problems: Testing for teaching (Chapter 4). Boston: Allyn and Bacon, Inc.

Weeks, E. (1979). Factors affecting the utilization of evaluation at the local level: An empirical test. In H. C. Schulberg and J. M. Jervell (Eds.), The evaluator and management (pp. 137-155). Beverly Hills, CA: Sage Publications.

Weiner, S. S., Rubin, D., & Sachse, T. (1977, August). Pathology in institutional structures for evaluation and a possible cure. Stanford Evaluation Consortium.

Weir, E. E. (1982). Causal understanding in naturalistic evaluation of educational programs. Unpublished doctoral dissertation, Ann Arbor, MI.

Weiss, C. H. (Ed.). (1977). Using social research in public policy making. Lexington, MA: Lexington-Heath.

Weiss, C. H. (1972). Utilization of evaluation: Toward comparative study. In C. H. Weiss (Ed.), Evaluating action programs: Readings in social action and education. Boston: Allyn and Bacon.

Wholey, J. (1980, April). Evaluability assessment as a tool for increasing the relevance of evaluation findings. Paper presented at the American Educational Research Association annual meeting, Boston, MA.

Wholey, J. S. (1977). Evaluability assessment. In L. Rutman (Ed.), Evaluation research methods: A basic guide. Beverly Hills, CA: Sage Publications.

Wilcox, B., & Bellamy, T. (1982). Design of high school programs for severely handicapped students. Baltimore: Paul H. Brookes Publishing Co.

Will, M. (1984). OSERS programming for the transition of youth with disabilities: Bridges for school to work life. Washington, DC: Office of Special Education and Rehabilitation Services.

Young, C. J. (1978, November). Evaluation utilization. Paper presented at the Evaluation Research Society annual meeting, Washington, DC.

Young, M. B. (1975). Evaluation and educational decision-making. Washington, DC: U.S. Department of Health, Education and Welfare, Office of Education.

Research Faculty at the University of Illinois

Janis Chadsey-Rusch
Visiting Assistant Professor
of Special Education

Lizanne DeStefano
Visiting Assistant Professor
of Educational Psychology

Jane Dowling
Visiting Assistant Professor
of Special Education

James W. Halle
Assistant Professor of
Special Education

Delwyn L. Harnisch
Assistant Professor of
Educational Psychology

Laird W. Heal
Professor of Special
Education

Robert L. Linn
Professor of Educational
Psychology

L. Allen Phelps
Professor of Vocational
Education

Adelle M. Renzaglia
Associate Professor of
Special Education

Frank R. Rusch
Professor of Special
Education

Richard P. Schutz
Visiting Assistant Professor
of Special Education

Robert E. Stake
Professor of Educational
Psychology

Institute Advisory Committee

Secondary Transition Intervention Effectiveness Institute
University of Illinois

Donn Brolin, Ph.D.
Department of Educational and Counseling Psychology
University of Missouri-Columbia

R. Brian Cobb, Ph.D.
Department of Special Education
University of Vermont

Marge Goldberg, *Co-Director*
Pacer Center
Minneapolis, Minnesota

Sally S. Höerr, *President*
National Parent Chain
Peoria, Illinois

Dean Inman, Ph.D.
Director of NERC
Center on Human Development
University of Oregon

Luanna Meyer, Ph.D.
Division of Special Education and Rehabilitation
Syracuse University

William Schill, Ph.D.
College of Education
University of Washington

Susan S. Suter, *Director*
Illinois Department of Rehabilitation Services

Edna Szymanski
American Rehabilitation Counseling Association
Senior Vocational Rehabilitation Counselor
New York State Office of Vocational Rehabilitation

Craig Thornton, Ph.D.
Mathematica Policy Research
Princeton, New Jersey

Paul Wehman, Ph.D.
Director
Rehabilitation, Research and Training Center
Virginia Commonwealth University

TRANSITION INSTITUTE AT ILLINOIS